A Completely Blind Video Integrity Oracle

Anish Mittal, Michele A. Saad, and Alan C. Bovik, *Fellow, IEEE*

***Abstract*— Considerable progress has been made toward developing still picture perceptual quality analyzers that do not require any reference picture and that are not trained on human opinion scores of distorted images. However, there do not yet exist any such completely blind video quality assessment (VQA) models. Here, we attempt to bridge this gap by developing a new VQA model called the video intrinsic integrity and distortion evaluation oracle (VIIDEO). The new model does not require the use of any additional information other than the video being quality evaluated. VIIDEO embodies models of intrinsic statistical regularities that are observed in natural vidoes, which are used to quantify disturbances introduced due to distortions. An algorithm derived from the VIIDEO model is thereby able to predict the quality of distorted videos without any external knowledge about the pristine source, anticipated distortions, or human judgments of video quality. Even with such a paucity of information, we are able to show that the VIIDEO algorithm performs much better than the legacy full reference quality measure MSE on the LIVE VQA database and delivers performance comparable with a leading human judgment trained blind VQA model. We believe that the VIIDEO algorithm is a significant step toward making real-time monitoring of completely blind video quality possible. The software release of VIIDEO can be obtained online (http://live.ece.utexas.edu/research/quality/VIIDEO_release.zip).**

***Index Terms*— Intrinsic video statistics, quality assessment, temporal self similarity, spatial domain.**

I. Introduction

OBJECTIVE VQA models seek to make predictions of visual quality automatically, in the absence of human judgments [1], [2]. Depending on the amount of available knowledge that they utilize, these models are commonly categorized as belonging to one of three different paradigms. Blind or NR VQA models assess the quality of videos without knowledge of the appearance of the video before it was distorted. At the other extreme are FR models, which require a 'clean', pristine reference video signal with respect to which the quality of the distorted video signal is assessed. Reduced-reference (RR) approaches lie somewhere between these extremes, utilizing some incomplete information regarding the reference image in addition to the distorted video [3].

Digital videos have become ubiquitous; already, more than 50% of both wireline and wireless data traffic is video data. Being able to monitor and control the perceptual quality of this traffic is a highly desirable goal that could be enabled by the development of 'completely blind' video quality analyzers that could be inserted into video networks or devices without any training or reference information [1]. Towards this end, we have developed and explain here a 'completely blind' video integrity oracle dubbed VIIDEO. Like the 'completely blind' picture quality analyzer NIQE [4], the approach taken here is simultaneously 'opinion-unaware' and 'distortion-unaware'. VIIDEO is even more penurious with respect to requiring exposure to other data: unlike NIQE, it is not trained on any data extracted from exemplar pristine videos, hence it is utterly 'content unaware' beyond using statistical models that can be shown to accurately characterize natural videos. While this may seem to be an extreme paucity of information, the use of perceptually relevant quantities yields results that are very promising. Indeed, the resulting algorithm predicts human judgments of video quality better than the long-standing full reference MSE on the LIVE VQA database [5].

This new NR VQA approach is derived based on intrinsic statistical regularities that are observed in natural videos. Deviations from these regularities alter their visual impression. Quantifying measurements of regularity (or lack thereof) under a natural video statistic model makes it possible to develop a 'quality analyzer' that can predict the visual quality of a distorted video without external knowledge of any kind beyond the underlying model.

Like the 'completely blind' picture quality analyzer in NIQE [4], VIIDEO does not require any distortion knowledge in the form of exemplar training videos containing anticipated distortions, or human opinions of video quality, distorted or otherwise. This is a significant advantage, given that creating VQA databases containing distorted videos and conducting large-scale studies to acquire co-registered human opinion scores is more involved than the creation of IQA (Image Quality Assessment) databases [5]–[7].

Accurately modeling the statistical regularities observed in natural videos is an important first step in understanding the perception of video quality by humans, yet it is very challenging. In the past we have proposed the use of exemplar natural pictures to serve as ground truth relative to which statistical regularities may be modeled [4]. Such an approach, although much more general than existing blind IQA models, is still limited in that it can only capture the common baseline characteristics of a specific collection of non-distorted content, and therefore may fail to represent some video specific intrinsic characteristics. Also, the construction of such a database requires the unbiased selection and maintenance of hundreds

Manuscript received August 12, 2014; revised March 6, 2015, August 30, 2015, and November 16, 2015; accepted November 16, 2015. Date of publication November 20, 2015; date of current version December 9, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Béatrice Pesquet-Popescu.

The authors are with the HERE, Intel, and also with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: mittal.anish@gmail.com; michele.saad@gmail.com; bovik@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2502725

of natural undistorted videos. This also raises the question of how many exemplar videos are needed to design an accurate natural video model, and how diverse and distinctive these need to be relative to each other and to the world of videos. Finally, given current limitations of image/video camera capture, distortions are inevitably introduced in the acquisition process and hence the procurement of perfectly natural 'pristine' videos is practically impossible.

In this article, we explain our 'quality aware' natural video statistics model in the space-time domain and describe the relevant temporal features that are derived from it and used to model inter subband correlations over local and global time spans. The overall model, which we call VIIDEO, is the basis of a VIIDEO algorithm that predicts video quality in a manner that correlates quite well with human judgments of video quality. We compare the performance of VIIDEO against existing state-of-the-art FR and NR VQA approaches. Before we describe the statistical and perceptual underpinnings of the VIIDEO model in detail, we review relevant prior work in the area of VQA.

II. Previous Work

The topic of NR VQA has been extensively studied and surveyed [8]–[17]. Unlike our problem here, we only conduct a brief review of recent relevant progress in the area. Nearly all prior NR VQA models have been 'distortion specific', meaning they were designed to predict the effect of a specific type of distortion on perceived quality. For example, Tan and Ghanbari [18], Vlachos [19], Suthaharan [20], and Muijs and Kirenko [21] proposed methods to assess blocking severity in distorted videos.

Methods for assessing multiple coincident distortion types have also been contemplated. Caviedes and Oberti compute a set of blocking, blurring, and sharpness features [22], Babu *et. al.* calculate a measure of blocking and packet-loss [23] and Farias and Mitra [24] measures blockiness, blur and noise. Massidda *et. al.* also propose a perception-based three-pronged NR metric for video distortion detection in 2.5G/3G systems which measures blockiness, bluriness and motion artifacts [17]. Dosselmann and Yang also estimate quality by measuring three types of impairments - noise, blocking and bit-error based color impairments [16]. Yang *et. al.* proposed an NR VQA algorithm that measures spatial distortion between a video block under consideration and its motion compensated block in the previous frame, where temporal distortion is computed as a function of the mean of the motion vectors [15]. Kawayokeita and Horita propose a model for NR VQA comprised of a frame quality measure and correction, asymmetric tracking and mean value filtering [14].

Other impairment specific NR VQA models include that proposed by Yang *et. al.*, where dropping severity is measured as a function of the number of frames dropped using timestamp information extracted from the video stream [13]. Lu proposed a method to measure blur caused by video compression [25], Pastrana-Vidal and Gicquel proposed an algorithm to measure frame-drops [12], Yamada *et. al.* proposed an algorithm to measure error-concealment effectiveness [11] and Naccari *et. al.* proposed a model for channel induced distortion for H.264/AVC coded videos [10]. Keimel *et. al.* proposed an NR VQA algorithm specifically for compressed HD videos [26] whereas Ong *et. al.* proposed a measure to monitor the quality of streamed videos by modeling the jerkiness between frames [9]. Saad and Bovik recently proposed a spatio-temporal natural scene statistics (NSS) model in the DCT domain that predicts the perceptual severity of MPEG-2, H.264, and two types of packet loss [27]. A linear kernel support vector regressor trained on a video database co-registered with human judgments is used to perform video quality prediction [28]. This model is generalizable for other distortion types but requires training on distorted videos and human opinion scores of them.

All of the blind VQA algorithms require knowledge of (one of possibly multiple) distortion types, or of the artifacts introduced by them, or of human opinion scores of exemplar distorted images. There exists no blind VQA model to date that can predict video quality in the absence of such strong *a priori* information.

III. Video Integrity Oracle

Towards reducing the performance dependency of visual quality prediction models on distortion type, human opinion or video content, we have developed a 'completely blind' VQA model that is based on a set of perceptually relevant temporal video statistic models of video frame difference signals. Algorithms that measure departures from these regularities have been used to capture distortion-induced anomalous behavior and to make visual quality inferences [3], [8]. By exploiting a greater variety of space-time statistical regularities, and by accounting for observable intrinsic properties of space-time band pass video correlations that are perceptually significant, we are able to develop a competitive VQA model that relies only on measuring perturbations of these properties.

A. Spatial Domain Natural Video Statistics

Our quality analyzer model derives insights from the reduced-reference VQA model proposed in [3] where it was shown that the bandpass filter coefficients of frame differences capture temporal statistical regularities arising from structures such as moving edges [3]. The ST-RRED model described there models the empirical distributions of the coefficients of multiscale wavelet transforms of frame differences. In a similar manner, but also inspired by the success of a recently proposed fast spatial domain IQA model [4], we instead analyse the local statistics of frame differences $\Delta F^{(t)}$ of videos that have been debiased and normalized using a different band pass perceptual model.

Define the frame difference ΔF^t between consecutive frames F^{2t+1} and F^{2t} of spatial dimensions $M \times N$ as follows:

$$\Delta F^t = F^{2t+1} - F^{2t} \quad \forall\, t \in \{0, 1, 2 \ldots \frac{(T-1)}{2}\} \qquad (1)$$

where T is the total number of frames.

The frame differences are operated on via processes of local mean removal and divisive contrast normalization following

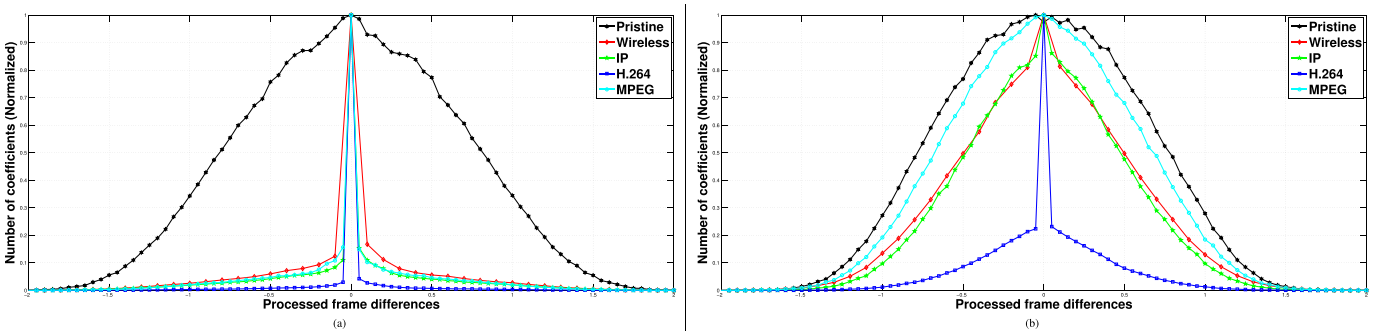

Fig. 1. Histograms of processed frame differences (2) computed on (a) 'mobile calendar' and (b) 'river bed' sequence and various distorted versions of these videos. All the videos are from the LIVE VQA database, which contains Wireless, IP, H.264 and MPEG distortions described in Section V [5].

the NSS model [29]:

$$\Delta\hat{F}^t(i,j) = \frac{\Delta F^t(i,j) - \mu^t(i,j)}{\sigma^t(i,j) + C} \tag{2}$$

over spatial indices $i \in \{1, 2 \dots M\}$, $j \in \{1, 2 \dots N\}$, over a set of consecutive frame time samples $t \in \{0, 1, 2 \dots \frac{(T-1)}{2}\}$ where

$$\mu^t(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} \Delta F^t(i+k, j+l) \tag{3}$$

and

$$\sigma^t(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} [\Delta F^t(i+k, j+l) - \mu(i,j)]^2} \tag{4}$$

estimate the local time-differenced luminance average and contrast, respectively, and where $w_{k,l} | k = -K, \dots, K$, $l = -L, \dots L$ is a Gaussian weighting function sampled out to 3 standard deviations and rescaled to unit volume. In our implementation, K = L = 3. $C = 1$ acts as a semi-saturation constant in the divisive normalization [30]. It prevents instabilities from occurring when the denominator tends to zero (eg., in the case of a temporal difference computed on a patch having a static background).

The space-time bandpass filtering and local nonlinear processing in (1)-(4) affect decorrelation of the video signal in a manner similar to neurons along the retino-cortical pathway [1].

B. Characterization of Patches

When computed from good quality, naturalistic still pictures (rather than frame differences), the coefficients (2) have been observed to reliably follow a generalized Gaussian distribution law. We have also found that debiased and normalized frame differences (2) are strongly decorrelated and Gaussianized, in agreement with the study on frame difference wavelet coefficients in [3]. Furthermore, when natural video frame differences are subjected to commonly encountered unnatural distortions, their processed coefficients no longer tend towards Gaussianity.

Fig. 1 plots the histogram of coefficients (2) computed on frame difference signals from the 'mobile calendar' and 'river bed' sequences and various distorted versions of them [5]. The histograms of the frame difference signals exhibit Gaussian-like appearance, whereas each distortion modifies the histograms in a characteristic way. For example, H.264 creates histograms having a more Laplacian-like appearance.

While transformed frame-difference coefficients are definitely more homogeneous for pristine frame-differences, the signs of adjacent coefficients also exhibit a regular structure, which is disturbed by the presence of distortion [31]. These deviations can be effectively probed by analyzing the sample distributions of products of pairs of adjacent coefficients computed along horizontal, vertical and diagonal spatial orientations: $\Delta\hat{F}^t(i,j)\Delta\hat{F}^t(i,j+1)$, $\Delta\hat{F}^t(i,j)\Delta\hat{F}^t(i+1,j)$, $\Delta\hat{F}^t(i,j)\Delta\hat{F}^t(i+1,j+1)$ and $\Delta\hat{F}^t(i,j)\Delta\hat{F}^t(i+1,j-1)$ for $i \in \{1, 2 \dots M\}$ and $j \in \{1, 2 \dots N\}$ [31]. The products of neighboring coefficients have shown to be well-modeled as following a zero mode asymmetric generalized Gaussian distribution (AGGD) [32]:

$$f(x; \gamma, \beta_l, \beta_r) = \begin{cases} \dfrac{\gamma}{(\beta_l + \beta_r)\,\Gamma\left(\frac{1}{\gamma}\right)} \exp\left(-\left(\dfrac{-x}{\beta_l}\right)^{\gamma}\right) & \forall x \le 0 \\ \dfrac{\gamma}{(\beta_l + \beta_r)\,\Gamma\left(\frac{1}{\gamma}\right)} \exp\left(-\left(\dfrac{x}{\beta_r}\right)^{\gamma}\right) & \forall x \ge 0. \end{cases} \tag{5}$$

The parameters of the AGGD $(\gamma, \beta_l, \beta_r)$ can be efficiently estimated using the moment-matching based approach in [32]. Using these models, we characterize the local statistics of frame difference (pristine or distorted) as follows. Each coefficient map (2) is partitioned into $P \times Q$ patches to enable a localized analysis of distortions. The estimated AGGD parameters are extracted from each patch to form model-based feature vectors that characterize each patch. By extracting estimates along the four orientations, 12 patch parameters are arrived at. Denote these features by ϕ_k^t, $k \in \{1, 2 \dots 12\}$ and the overall feature vector by $\Phi_{x,y}^t$, $t \in \{1, 2 \dots (T-1)/2\}$, $(x, y) \in \{1, 2 \dots P\} \times \{1, 2 \dots Q\}$.

The coefficient products statistically modeled by (5) may be viewed as highly localized correlation measurements. As discussed in section IV, there is evidence that local correlation operations may be executed by visual neurons (fed by area V1)

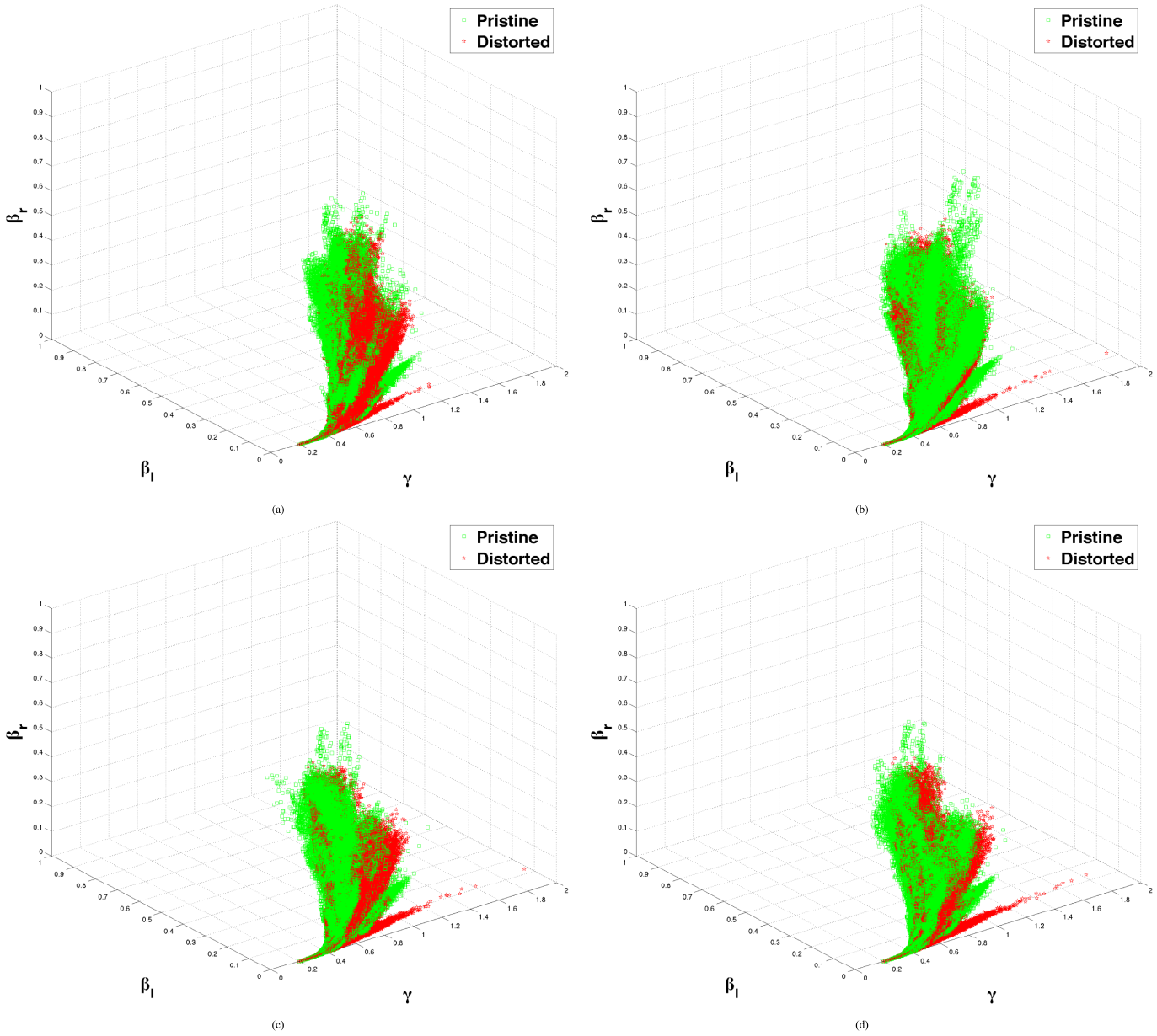

Fig. 2. 3D scatter plots between shape, left scale and right scale obtained by fitting the AGGD model to (a) horizontal spatial paired products, (b) vertical spatial paired products, (c) on-diagonal spatial paired products and (d) off-diagonal paired spatial products using all pristine and distorted videos from the LIVE VQA database [5].

involved in computing visual forms, which would be affected by distortion.

Fig. 2 shows 3D scatter plots between the shape, left scale and right scale parameters obtained by fitting AGGD to horizontal spatial paired products, vertical spatial paired products, on diagonal spatial paired products and off diagonal paired spatial products using pristine and distorted versions of all sequences together from the LIVE VQA database [5]. As may be observed from Fig. 2, there is substantial overlap between the features obtained from pristine and distorted videos in the scatter plot when features from all videos are plotted together. This suggests that the features by themselves are not content agnostic, hence further processing is required to reduce their content sensitivity.

Fig. 3 shows 3D scatter plot between temporal changes in shape, temporal changes in left scale and temporal changes in right scale obtained from horizontal spatial paired products, vertical spatial paired products, on-diagonal spatial paired products and off-diagonal paired spatial products using all of the pristine and distorted videos from the LIVE VQA database [5]. The figure strongly suggests that pristine and distorted videos have a much clearer separation in temporal feature change space. In the next section, we discuss how to utilize this useful information captured using temporal change statistics.

C. Inter Sub-Band Statistics

It is important to understand that variations in the local temporal statistics of a given space-time video patch capture could

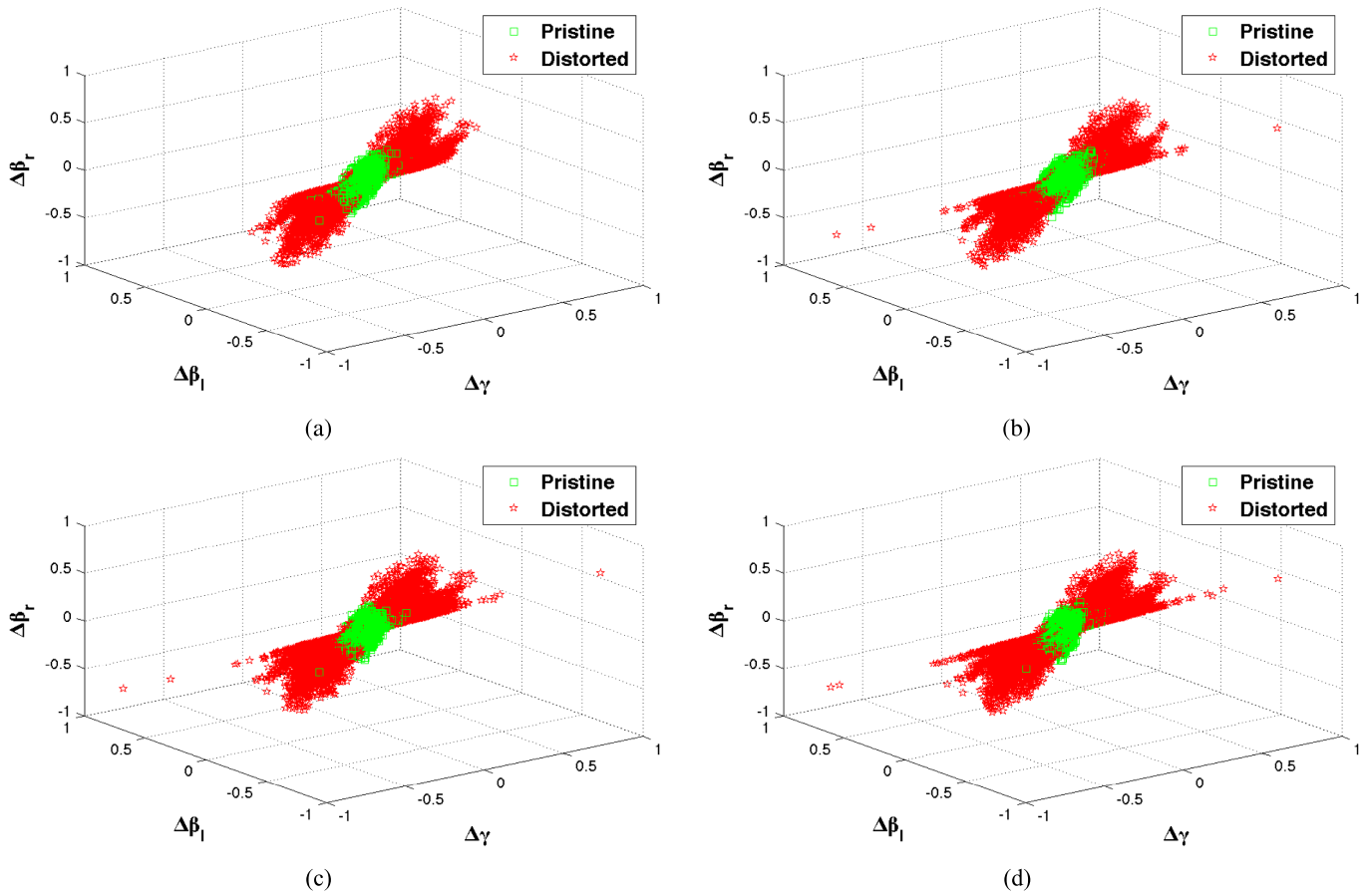

Fig. 3. 3D scatter plots between temporal change in shape parameter, temporal change in left scale parameter and temporal change in right scale parameter computed on (a) horizontal spatial paired products, (b) vertical spatial paired products, (c) on-diagonal spatial paired products and (d) off-diagonal paired spatial products using all pristine and distorted videos from the LIVE VQA database [5].

be indicative either of naturalistic motion or of temporal distortion. However, motion induced changes are associated with strong correlations between the transformed frame difference coefficients at fine and coarse scales over time, while temporal distortion induced changes tend to cause transient statistical aberrations. Towards capturing these differences, the VIIDEO model measures the temporal variations of statistical model feature correlations between different scales of transformed image coefficients in a local way.

As discussed in Section IV, there is evidence that integrated products of local visual signals are used in extra-cortical area V2 of the visual brain, and that these measurements are implicated in the early computation of visual shapes. However, the nature of these computations is not yet well understood, hence we take the simplest available path by computing local space-time empirical correlations (products) of measurements derived from the space-time distribution models (1)-(5).

Thus, first compute low pass filtered frame difference coefficients:

$$\Delta G^t(i,j) = \sum_{k=-K}^{K}\sum_{l=-L}^{L} w_{k,l}\,\Delta F^t(i+k, j+l) \tag{6}$$

where as before $t \in \{0,1,2\ldots\frac{(T-1)}{2}\}, i \in \{1,2\ldots M\}$, $j \in \{1,2\ldots N\}$ are spatial indices, M and N are the image dimensions, T is the total number of frames and $w_{k,l}$; $k = -K,\ldots,K, l = -L,\ldots L$ is a 2D circularly-symmetric Gaussian weighting function (the same as was used in Section III-A).

The feature vectors $\Gamma^t_{x,y}$ are computed from the low pass filtered frame differences $\Delta G^t(i,j)$, following the same procedure described earlier for computing the features $\Phi^t_{x,y}$ from the frame difference coefficients $\Delta F^t(i,j)$.

We model the change in statistics using temporal feature vector differences as follows:

$$\Delta\Phi^t_{x,y} = \Phi^{t+1}_{x,y} - \Phi^t_{x,y} \tag{7}$$

$$\Delta\Gamma^t_{x,y} = \Gamma^{t+1}_{x,y} - \Gamma^t_{x,y} \tag{8}$$

$\forall\ t \in \{1,2\ldots(T-1)/2\}$; and $\forall\ (x,y) \in \{1,2\ldots P\} \times \{1,2\ldots Q\}$.

For each time instant (frame) indexed t, these differences are captured or windowed over a span of frames of length S, denoted by the sets:

$$A^t_{x,y} = \{\Delta\Phi^{t+s}_{x,y} \quad \forall\ s \in \{1,2\ldots S\}\} \tag{9}$$

$$B^t_{x,y} = \{\Delta\Gamma^{t+s}_{x,y} \quad \forall\ s \in \{1,2\ldots S\}\} \tag{10}$$

where $A^t_{x,y}$ and $B^t_{x,y}$ are measured at every $t+\eta k$ where η is the stride and $k = 0,1,2\ldots$.

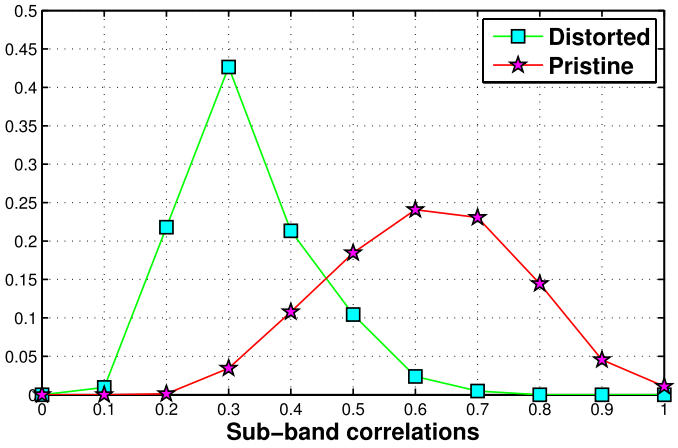

Fig. 4. Probability distributions depicting inter subband empirical correlations on undistorted and distorted natural videos in the LIVE VQA database [5].

The empirical correlation coefficient vector $\hat{\Theta}^{t+\eta k}$ is defined as $C[A^{t+\eta k}_{x,y}, B^{t+\eta k}_{x,y}]$ where $C[A_{x,y}, B_{x,y}]$ is the normalized sample covariance $A_{x,y} = \{a^1_{x,y}, a^2_{x,y} \ldots a^S_{x,y}\}$ and $B_{x,y} = \{b^1_{x,y}, b^2_{x,y} \ldots b^S_{x,y}\}$ defined separately for each feature in the vector:

$$C[A_{x,y}, B_{x,y}] = \sum_{x,y,s} \left(\frac{a^s_{x,y} - \mu\left(A_{x,y}\right)}{\sigma\left(A_{x,y}\right)} \right) \left(\frac{b^s_{x,y} - \mu\left(B_{x,y}\right)}{\sigma\left(B_{x,y}\right)} \right) \tag{11}$$

where $\mu\left(A_{x,y}\right) = \frac{\sum_{x,y,s} a^s_{x,y}}{PQS}$ and $\sigma\left(A_{x,y}\right) = \frac{\sqrt{\sum_{x,y,s}\left[a^s_{x,y} - \mu\left(A_{x,y}\right)\right]^2}}{PQS}$.

To reiterate, $\hat{\Theta^{t+\eta k}}$ represents the empirical correlation coefficient feature vector that contains the empirical correlation coefficients $\hat{\theta}^{t+\eta k}_f$ obtained on all features such that $f \in \{1, 2 \ldots 12\}$. To be clear, the 12 correlation coefficients are defined based on parameters computed using distribution fits to point-wise spatial products of transformed frame-difference coefficients (2).

Over each length S time span indexed $t + \eta k$, all the correlation coefficients are aggregated by averaging as follows:

$$\Omega^{t+\eta k} = \Sigma_f \hat{\theta}^{t+\eta k}_f \quad \forall\, f \in \{1, 2, 3 \ldots 12\} \tag{12}$$

Fig. 4 shows the histograms of Ω^t for both pristine and distorted videos from the LIVE VQA database [5]. As seen clearly in the Fig. 4, the sub band correlations are higher for pristine videos as compared to distorted videos. Although there is some overlap between the histograms, they are generally well separated.

The final score is obtained as follows:

$$\Psi = \Sigma_{t+\eta k} \Omega^{t+\eta k} \tag{13}$$

thereby capturing both average (summed) statistical measurements of video quality.

IV. Relevant Perceptual Mechanisms

More detailed discussion of the first and second order band pass statistical models just described, and their possible relationship to visual processing architectures is in order. The processed frame difference coefficients (2) represent a simple realization of center-surround band pass processing by the earliest post-retinal visual neurons, which accomplishes significant reductions in spatial redundancy. The model also includes temporal decorrelation by frame differencing and divisive normalization by neighboring energy responses, which serves to further spatially decorrelate and gaussianize the visual signal. These processes broadly mimick temporal lag decorrelation [33] and adaptive gain control further along the visual pathway, and extending into area V1 (primary cortex) [30]. Notably, the debiased and normalized space-time signal is not orientation selective, unlike the responses of visual cortical neurons [34], [35]. However, in the context of visual quality assessment, orientation selectivity has not yet been shown to significantly improve predictions of the perceptual impact of statistical impairments as can be seen, for example by the favorable performance of the spatially isotropic model [31] relative to the orientation-selective DIIVINE index [36].

The statistical model (5) of empirical pairwise products or correlations of the perceptually processed signal (2) also has a possible interpretation with respect to processing further along the visual pathway. In particular, the neurons in area V2 of primate cortex, which is driven by an extensive projection of V1 responses, is implicated in the representation of higher-order spatio-temporal forms such as contours, angles and disparities. A recent comprehensive study of the population of V1 neurons that project to V2 [37] showed that, in addition to constituting the largest projection of V1 responses, it also includes the gamut of V1 functionality, including directional and absolute motion sensitivity, orientation sensitivity and multi-scale. This suggests that the residual correlations present in the cortical signal may be processed in V2 towards the early construction of space-time forms, which portends a relevance to both ventral and dorsal cortical processing, both of which are critical to the perception of video quality. This is supported by measurements of V2 responses to complex shapes [38], combinations of orientations [39], and contour angles [40] which suggest a functional integration of V1 responses. Notably, both the preponderance of V1 neurons projecting to V2 [37], as well as V2 neurons [41] exhibit strong surround suppression, which may contribute to selectivity to complex space-time forms. Currently, there is not any settled-on model of V2 form processing, although it appears to include local space-time integration of V1 responses [42], which broadly equates to the computation of local correlations and probing of correlated structures. It may be expected that correlations induced by (or destroyed by) distortion will deeply affect these computations. In the absence of an adequate model, and following Occam's Razor, we follow the simple strategy in [43] of computing space-time correlations (used successfully there to model visual crowding) as a simple measure of the information (possibly distorted) available to V2 units.

V. Quality-Feature Goodness Tests

We posit that the intrinsic statistics of a video test sample should satisfy three criteria, if they are to be useful for predicting perceived video quality. First, the statistics should be

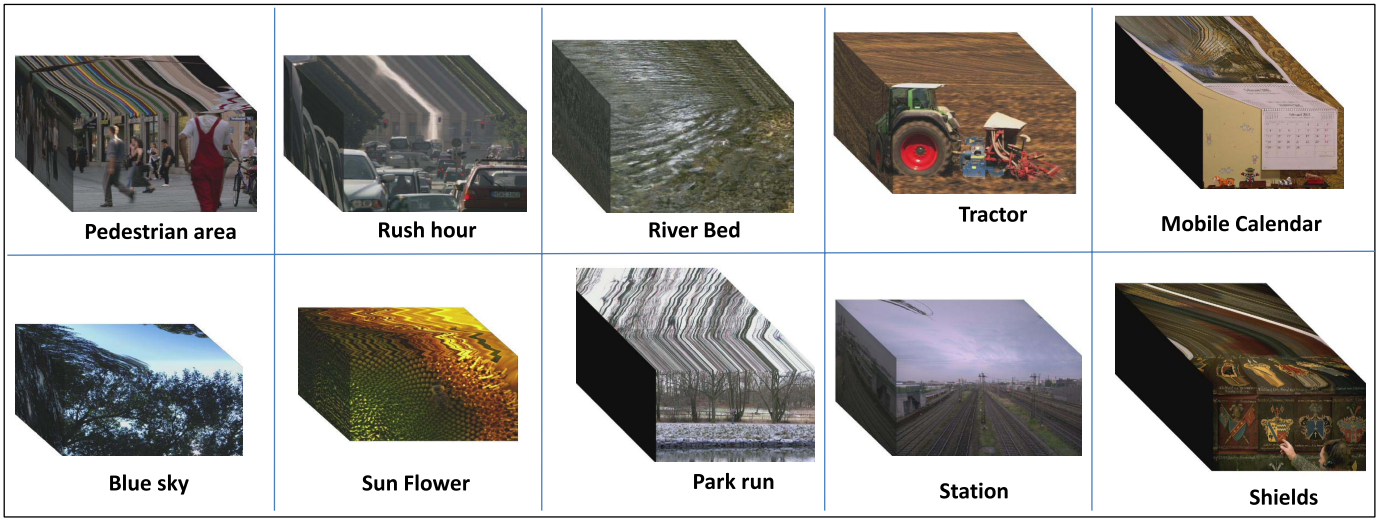

Fig. 5. Depiction of 10 reference video contents on which different kinds of distortions with varying degrees were introduced to make the LIVE VQA database.

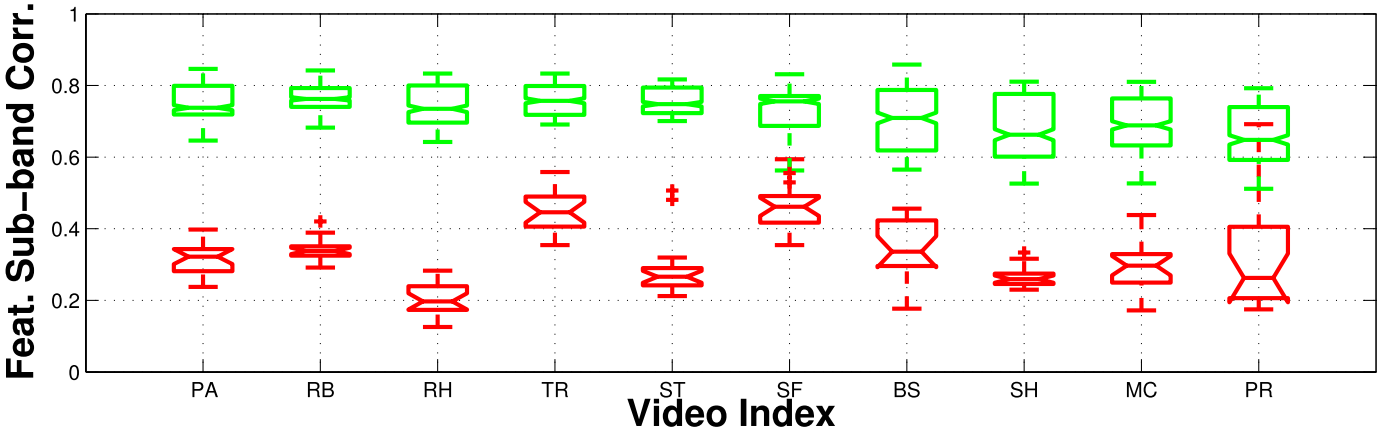

Fig. 6. Box plot depicting inter subband correlations on undistorted and distorted natural videos in the LIVE VQA database [5] for the average of the features. Since the inter subband correlations are higher for pristine videos as compared to their distorted versions, boxes at the top of the figures (in green) are obtained from pristine videos whereas the red boxes are obtained from distorted videos.

regular over pristine videos i.e., they should reliably follow a model. Second, the regularity should be destroyed on distorted video samples. Third, the loss of regularity should vary with the degree of perceived distortion in the distorted video.

We use the videos from the LIVE Video Quality Assessment Database [5] to conduct the tests. The LIVE Video Quality Assessment Database contains 10 reference videos, depicted in Fig. 5 with frames of size 432 × 768 - *'pedestrian area'*, *'river bed'*, *'rush hour'*, *'tractor'*, *'station'*, *'sun flower'* contain 250 frames at 25 fps, *'blue sky'* contains 217 frames at 25 fps, *'shields'*, *'mobile calender'*, *'park run'* contain 500 frames at 50 fps. Derived from these pristine videos is a set of 150 distorted videos with a span of 4 distortion categories, including compression artifacts due to MPEG and H.264, errors induced by transmission over IP networks and errors introduced due to transmission over wireless networks.

Currently, LIVE is as general as any video quality database in terms of distortion diversity. We greatly desire to be able to test on much more distortion-generic video data, but our own efforts in this direction are many months or years away from fruition.

A. Comparing Pristine and Distorted Temporal Video Correlations

Real-world naturalistic videos exhibit statistical regularities and strong correlations over both space and time. As just discussed in Section IV, we hypothesize that the neural circuits that conduct spatio-temporal perception have adapted to these regularities and are tuned towards exploiting correlations that exist in space-time visual signals to accomplish a wide variety of high level perceptual tasks. Characterizing these regularities on 'pristine' videos not suffering from obvious distortion is an important step towards developing models of human sensitivity to visual impairments that destroy the intrinsic regularity and correlation structure of naturalistic videos.

Fig. 6 shows box plots of Ω^t for 10 different video contents drawn from the LIVE VQA database [5]. The boxes at the top of the figures (in green) derive from the LIVE pristine video

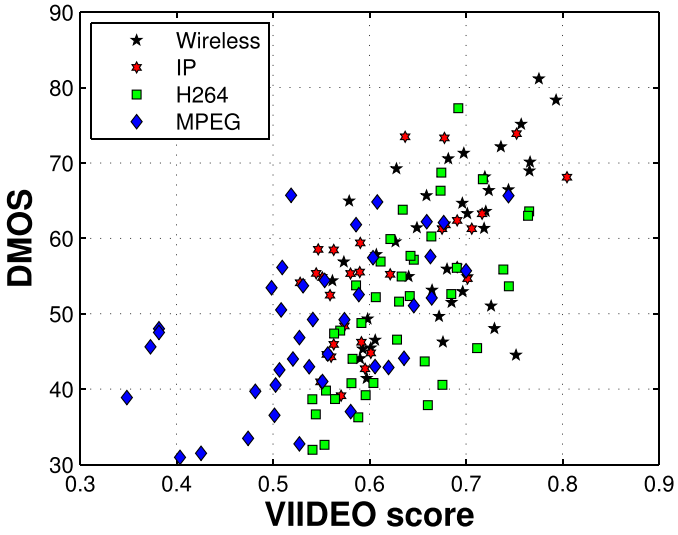

Fig. 7. Final quality predictions using VIIDEO on the LIVE VQA database [5].

content whereas the boxes at the bottom of each figure (in red) were obtained from all of the distorted videos present in the LIVE VQA database [5]. Clearly, the inter sub-band correlations are much higher on the pristine video contents than on the distorted contents. Moreover, the correlation features of the pristine videos are remarkably consistent, despite the significant diversity of content.

B. Perceptual Relevance of Temporal Correlation Features

Of course, the most important criteria that a perceptually motivated intrinsic video statistic should satisfy in this context is that it correlates highly with human judgments of visual quality. To test this relevance, we plotted the VIIDEO score Ψ against the DMOS of all 150 distorted videos in the LIVE database, shown in Fig. 7. By examining the scatter plots, it may be seen that the perceptual NSS based VIIDEO score presents an approximately linear (aggregate) trend against subjective judgments. We only report test results on the distorted videos since DMOS $= 0$ for all pristine videos, which is a level of consistency beyond any human judgment. Testing on pristine videos greatly boosts the numerical performance of VIIDEO (even against other models) but is not as accurate a comparison.

VI. Experiments

To test the VIIDEO features, the patch size parameters P and Q were set to 72, and the analysis interval length S and temporal stride η were set to 1 and 0.5 second respectively.

A. Algorithm Prediction Performance

We used SROCC, and Pearson's (linear) correlation coefficient (LCC) to test the model. The VIIDEO scores were passed through a logistic non-linearity [44] to map to DMOS before computing LCC.

Since it is of value to study the outlier cases on which VIIDEO delivers worst performance, we recorded the errors incurred while fitting the non-linearity on the complete data set during this procedure. The worst 25% of all performers in terms of score prediction were bucketed based on the content from which they were generated. The results are tabulated in Table I. VIIDEO delivered the worst performance on 'Shields' and 'Mobile Calendar'. These two videos present extremes of motion (slow motion translation) along with a high degree of spatial activity.

We next compared VIIDEO with five FR indices: MSE, MS-SSIM [45], MOVIE [46], VQM [47], ST-MAD [48], one RR index: STRRED [3] and one learning based blind index: V-BLIINDS [8]. We report the performance of FR and RR indices on the entire database including pristine and distorted videos in Tables II and III.

Although there exist differences in the SROCC and LCC correlations between the different algorithms (see Table II), it is important to determine whether these differences are not statistically relevant. To evaluate the statistical significance

TABLE I

Outlier Cases on Which VIIDEO Delivers Worst Performance, Errors Were Recorded While Passing VIIDEO Scores Through the Logistic Non-Linearity [44] to Map to the DMOS on the Complete Data Set. Worst 25% Videos in Terms of Score Prediction Were Bucketed Based on the Content Using Which They Were Generated

Content Name	Misclassification Rate
Pedestrian Area	3/15
River Bed	4/15
Rush Hour	4/15
Tractor	5/15
Station	2/15
Sun flower	1/15
Blue sky	5/15
Shields	6/15
Mobile Calendar	6/15
Park Run	2/15
Total	38/150

TABLE II

SROCC of Different VQA Algorithms Against DMOS on LIVE VQA Database

Algorithm	Wireless	IP	H.264	MPEG	Overall
MSE	0.657	0.417	0.458	0.386	0.540
MS-SSIM	0.729	0.653	0.731	0.668	0.736
MOVIE	**0.811**	0.716	0.766	0.773	0.789
VQM	0.721	0.638	0.652	0.781	0.702
STMAD	0.810	**0.776**	**0.902**	**0.846**	**0.825**
STRRED	0.786	0.771	0.820	0.720	0.802
VIIDEO	0.532	0.612	0.674	0.556	0.624

TABLE III

LCC of Different VQA Algorithms Against DMOS on LIVE VQA Database

Algorithm	Wireless	IP	H.264	MPEG	Overall
MSE	0.684	0.525	0.591	0.420	0.563
MS-SSIM	0.718	0.776	0.742	0.622	0.747
MOVIE	0.842	0.766	0.814	0.798	0.813
VQM	0.755	0.667	0.666	0.813	0.730
STMAD	**0.859**	0.806	**0.915**	**0.856**	**0.833**
STRRED	0.806	**0.815**	0.823	0.758	0.812
VIIDEO	0.625	0.737	0.692	0.631	0.651

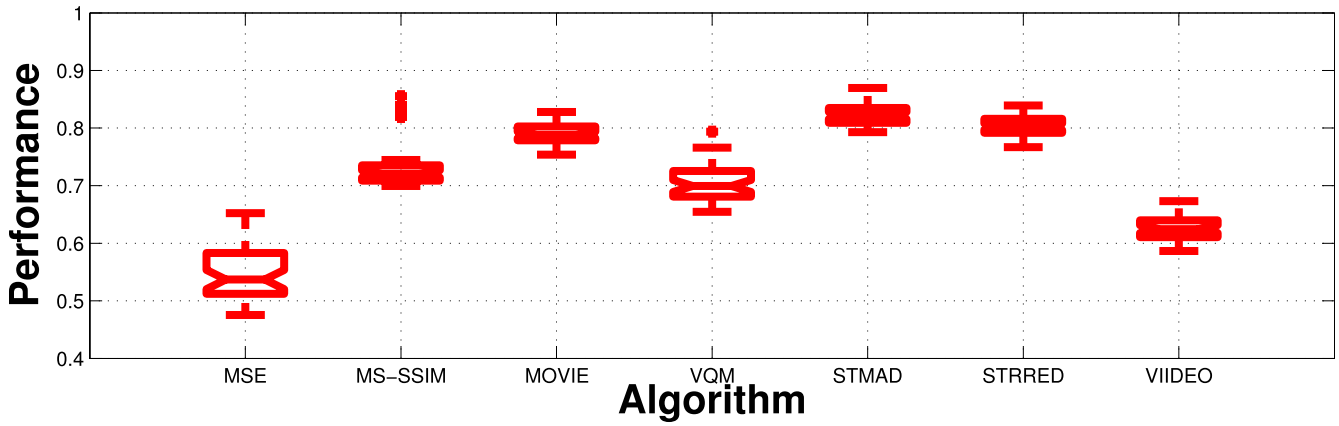

Fig. 8. Box plot of SROCC scores across 45 trials where every trial involved randomly picking 8 out of 10 video contents (including the pristine and distorted versions) in LIVE VQA database [5].

TABLE IV

RESULTS OF ONE SIDED t-TEST PERFORMED BETWEEN SROCC VALUES OF VARIOUS FULL REFERENCE VQA ALGORITHMS. A VALUE OF '1' INDICATES THAT THE ROW ALGORITHM IS STATISTICALLY SUPERIOR TO THE COLUMN ALGORITHM; '−1' INDICATES THAT THE ROW IS WORSE THAN THE COLUMN; A VALUE OF '0' INDICATES THAT THE TWO ALGORITHMS ARE STATISTICALLY INDISTINGUISHABLE

Algorithm	MSE	MS-SSIM	MOVIE	VQM	STMAD	STRRED	VIIDEO
MSE	0	-1	-1	-1	-1	-1	-1
MS-SSIM	1	0	-1	1	-1	-1	1
MOVIE	1	1	0	1	-1	-1	1
VQM	1	-1	-1	0	-1	-1	1
STMAD	1	1	1	1	0	1	1
STRRED	1	1	1	1	-1	0	1
VIIDEO	1	-1	-1	-1	-1	-1	0

of performance of each of the algorithms considered, we performed the t-test [44] on the SROCC values obtained from the 45 trials, where every trial involved randomly picking 8 of 10 video contents (including the pristine and distorted versions) in the LIVE VQA database [5]. Figure 8 shows the box plot. We tabulated the results in Table IV. The null hypothesis is that the mean correlation for the (row) algorithm is equal to the mean correlation for the (column) algorithm with a confidence of 95%. The alternate hypothesis is that the mean correlation of the row is greater than or lesser than the mean correlation of the column. A value of '1' in the table indicates that the row algorithm is statistically superior to the column algorithm, whereas a '−1' indicates that the row is statistically worse than the column. A value of '0' indicates that the row and column are statistically indistinguishable (or equivalent), i.e., we could not reject the null hypothesis at the 95% confidence level.

As is evident from the results, VIIDEO correlates better with human judgments of visual quality than the full reference measure MSE, which is remarkable given that VIIDEO does not incorporate or depend on any kind of external information. Unsurprisingly, VIIDEO remains inferior to the perceptually relevant full reference MS-SSIM and MOVIE, but this also indicates that there may still be room for improvement.

The only high performance general-purpose blind VQA algorithm available to compare VIIDEO with is

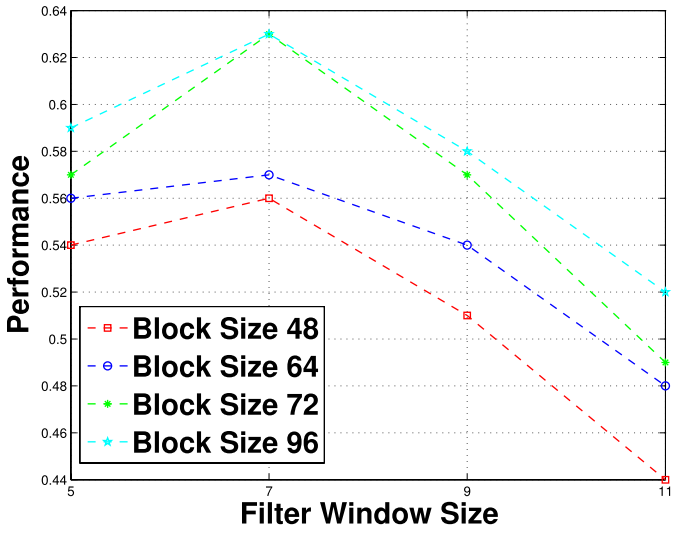

Fig. 9. SROCC performance of VIIDEO index against different filter window sizes and block sizes.

TABLE V

MEDIAN SROCC OF DIFFERENT BLIND VQA ALGORITHMS AGAINST DMOS ON EVERY POSSIBLE COMBINATION OF TRAIN/TEST SET SPLITS. 80% OF CONTENT USED FOR TRAINING ON LIVE VQA DATABASE

Algorithm	Wireless	IP	H.264	MPEG	Overall
V-BLIINDS	**0.691**	**0.600**	0.643	**0.667**	**0.735**
VIIDEO	0.548	**0.600**	**0.762**	0.571	0.651

the learning based V-BLIINDS model [27], which requires a training procedure to calibrate the regressor module. Hence to make this comparison, we repeatedly divided the LIVE VQA database into randomly chosen 80% training and 20% testing subsets - taking care that no overlap occurs between train and test content. This train-test procedure was done on every possible combination of train/test set splits to minimize any bias due to the video content used for training. We report the median performance across all iterations in Tables V and VI. As may be observed, V-BLIINDS performs at a higher level than VIIDEO, which is not surprising given that it

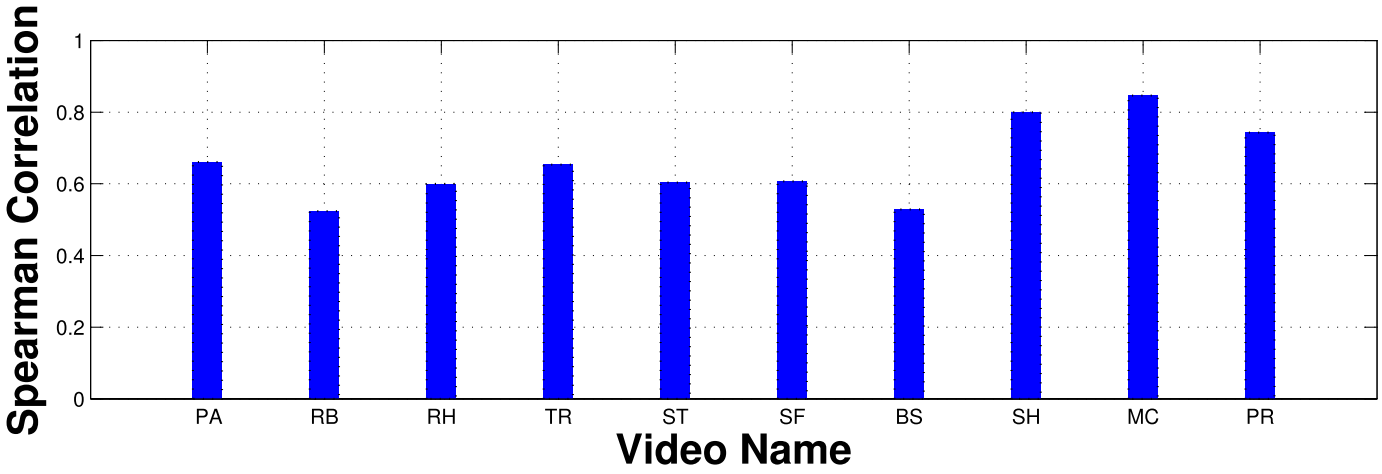

Fig. 10. Performance on the test videos *'pedestrian area', 'river bed', 'rush hour', 'tractor', 'station', 'sun flower', 'blue sky', 'shields', 'mobile calender', 'park run'* taken one at a time.

TABLE VI

MEDIAN LCC OF DIFFERENT BLIND VQA ALGORITHMS AGAINST DMOS ON EVERY POSSIBLE COMBINATION OF TRAIN/TEST SET SPLITS. 80% OF CONTENT USED FOR TRAINING ON LIVE VQA DATABASE

Algorithm	Wireless	IP	H.264	MPEG	Overall
V- BLIINDS	**0.844**	**0.852**	**0.956**	**0.949**	**0.790**
VIIDEO	0.740	0.848	0.886	0.872	0.701

has the advantage of distortion specific calibration during training. The performance numbers reported for VIIDEO in Tables III and II against Tables VI and V are some what different. However, the test sets used to evaluate VIIDEO in these separate analysis differ. The test set used to compute the performance numbers reported in Tables II and III contain distorted videos generated from all 10 contents present in the LIVE Video Quality Assessment Database [5], the test data used to compute the performance numbers reported in Tables V and VI contain distorted videos generated from only 2 video contents at a time. Therefore, there is better alignment between predicted scores in the latter case, since the variation in content across videos is smaller and hence higher performance may be observed. However, since the algorithm being compared is learning based, it does require other distorted videos with co-registered human judgments to be trained on. This is to some degree, a reflection of the limited size of all available databases.

B. Parameter Variation

We tested VIIDEO using different filter window sizes and block sizes to explore the dependence of the algorithm on these parameters. Figure 9 shows the SROCC performance plotted against filter window sizes and block sizes. As may be observed from the figure, the algorithm performed best for intermediate block size [64 72] with Gaussian filter of size 7.

C. Linearity

Linearity refers to a the relationship when the two variables can be represented as directly proportional to each other and

TABLE VII

LINEARITY OF DIFFERENT VQA ALGORITHMS AGAINST DMOS ON LIVE VQA DATABASE

Algorithm	Wireless	IP	H.264	MPEG	Overall
MSE	0.633	0.436	0.571	0.364	0.541
MS-SSIM	0.682	0.685	0.584	0.622	0.680
MOVIE	**0.836**	0.759	0.787	0.725	0.795
VQM	0.734	0.649	0.628	0.772	0.714
STMAD	0.802	**0.796**	**0.908**	**0.820**	**0.823**
STRRED	0.766	0.718	0.765	0.744	0.693
VIIDEO	0.556	0.662	0.653	0.556	0.631

can be graphically represented as following a straight line. We measure the degree of linearity using LCC. It is important to notice here that unlike the results reported in Tables III and VI, where the algorithms are passed through a logistic non linearity, we directly find the coefficient of linear correlation between the algorithm score and the human judgments.

Linearity is a highly desirable property of a VQA index since it makes the model more tractable and amenable for its use as a quality optimization function in video processing applications. As may be observed from Table VII, the performance of VIIDEO exceeds that of the MSE in terms of linearity as well.

D. Content Dependence

We also explored the way that VIIDEO behaves across diverse video contents. Fig. 10 shows the performance of VIIDEO on each of the video content present in the LIVE VQA database. Although the index can be observed to perform slightly better on some videos than others, the variation in performance lay within fairly small range.

VII. DISCUSSION

We have proposed a 'completely blind' natural video statistics based quality assessment model - Video Intrinsic Integrity and Distortion Evaluation Oracle (VIIDEO). It does not model any distortion specific information, but only models the statistical 'naturalness' (or lack thereof) of the video. We described

how the inter subband correlations can be used to quantify the degree of distortion present in the video and hence to predict human judgments of video quality.

We also analyzed the time complexity of every step in the VIIDEO algorithm. The filtering and divisive normalization operations are the most computationally expensive steps, with complexity $O(TMN\log_2(NM))$. However, since both of the steps involve point-based pixel wise computations, they are quite parallel in nature and can easily achieve linear scaling with the number of processors deployed to achieve the task.

We also undertook a thorough evaluation of the VIIDEO model in terms of the correlation of the quality predictions it makes with human judgments, and demonstrated that VIIDEO performs better in this regard than the FR MSE metric.

There is still scope for improvement by incorporating better models of motion for integration into blind VQA algorithms. This may include more complete modeling of temporal filtering in the lateral geniculate nucleus (LGN) and motion processing in Areas MT/V5 and MST of extrastriate cortex [49]–[51]. The development of more detailed models of functional processing in cortical area V2 remains a very energetic research area, with obvious positive implications for applied visual neuroscience problems of this kind.

Future work could involve faster implementation of the algorithm for real time video monitoring applications. We envision that the proposed VIIDEO model could help solve the resource allocation problem, by modeling the quality of video traffic with the intent of optimizing rate control protocols that heighten the end-user's perceptual experience. This would be facilitated by existing databases of rate-switched videos, e.g. [7].

REFERENCES

[1] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.

[2] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.

[3] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 2012.

[4] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2012.

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

[6] F. de Simone, M. Naccari, M. Tagliasacchi, F. Dufaux, S. Tubaro, and T. Ebrahimi, "Subjective assessment of H.264/AVC video sequences transmitted over a noisy channel," in *Proc. QoMEX*, Jul. 2009, pp. 204–209.

[7] 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 J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 652–671, Oct. 2012.

[8] M. A. Saad and A. C. Bovik, "Blind quality assessment of natural videos using motion coherency," in *Proc. IEEE Asilomar Conf. Signals, Syst., Comput.*, Nov. 2012, pp. 332–336.

[9] E. P. Ong *et al.*, "Video quality monitoring of streamed videos," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 1153–1156.

[10] M. Naccari, M. Tagliasacchi, F. Pereira, and S. Tubaro, "No-reference modeling of the channel induced distortion at the decoder for H.264/AVC video coding," in *Proc. 15th IEEE Int. Conf. Image Process.*, Oct. 2008, pp. 2324–2327.

[11] T. Yamada, Y. Miyamoto, and M. Serizawa, "No-reference video quality estimation based on error-concealment effectiveness," in *Proc. Packet Video*, Nov. 2007, pp. 288–293.

[12] R. R. Pastrana-Vidal and J.-C. Gicquel, "Automatic quality assessment of video fluidity impairments using a no-reference metric," in *Proc. Int. Workshop Video Process. Quality Metrics Consumer Electron.*, 2006, pp. 1–6.

[13] K.-C. Yang, C. C. Guest, K. El-Maleh, and P. K. Das, "Perceptual temporal quality metric for compressed video," *IEEE Trans. Multimedia*, vol. 9, no. 7, pp. 1528–1535, Nov. 2007.

[14] Y. Kawayokeita and Y. Horita, "NR objective continuous video quality assessment model based on frame quality measure," in *Proc. IEEE Int. Conf. Image Process.*, San Diego, CA, USA, Oct. 2008, pp. 385–388.

[15] F. Yang, S. Wan, Y. Chang, and H. R. Wu, "A novel objective no-reference metric for digital video quality assessment," *IEEE Signal Process. Lett.*, vol. 12, no. 10, pp. 685–688, Oct. 2005.

[16] R. Dosselmann and X. D. Yang, "A prototype no-reference video quality system," in *Proc. 4th Can. Conf. Comput. Robot Vis.*, May 2007, pp. 411–417.

[17] F. Massidda, D. D. Giusto, and C. Perra, "No reference video quality estimation based on human visual system for 2.5/3G devices," *Proc. SPIE*, vol. 5666, pp. 168–179, Mar. 2005.

[18] K. T. Tan and M. Ghanbari, "Blockiness detection for MPEG2-coded video," *IEEE Signal Process. Lett.*, vol. 7, no. 8, pp. 213–215, Aug. 2000.

[19] T. Vlachos, "Detection of blocking artifacts in compressed video," *Electron. Lett.*, vol. 36, no. 13, pp. 1106–1108, Jun. 2000.

[20] S. Suthaharan, "Perceptual quality metric for digital video coding," *Electron. Lett.*, vol. 39, no. 5, pp. 431–433, 2003.

[21] R. Muijs and I. Kirenko, "A no-reference blocking artifact measure for adaptive video processing," in *Proc. Eur. Signal Process. Conf.*, 2005, pp. 1–4.

[22] J. E. Caviedes and F. Oberti, "No-reference quality metric for degraded and enhanced video," *Proc. SPIE*, vol. 5150, pp. 621–632, Jun. 2003.

[23] R. V. Babu, A. S. Bopardikar, A. Perkis, and O. I. Hillestad, "No-reference metrics for video streaming applications," in *Proc. Int. Workshop Packet Video*, 2004.

[24] M. C. Q. Farias and S. K. Mitra, "No-reference video quality metric based on artifact measurements," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3. Genoa, Italy, 2005, p. III-141.

[25] J. Lu, "Image analysis for video artifact estimation and measurement," *Proc. SPIE*, vol. 4301, pp. 166–174, Apr. 2001.

[26] C. Keimel, T. Oelbaum, and K. Diepold, "No-reference video quality evaluation for high-definition video," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 1145–1148.

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

[28] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "kernlab—An S4 package for kernel methods in R," *J. Statist. Softw.*, vol. 11, no. 9, pp. 1–20, 2004.

[29] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[30] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *J. Neurosci.*, vol. 9, no. 2, pp. 181–197, 1992.

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

[32] N.-E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *Proc. IEEE Int. Conf. Image Process.*, Nov. 2009, pp. 2281–2284.

[33] D. W. Dong and J. J. Atick, "Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus," *Netw., Comput. Neural Syst.*, vol. 6, no. 2, pp. 159–178, 1995.

[34] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A*, vol. 2, no. 7, pp. 1160–1169, 1985.

[35] M. Clark and A. C. Bovik, "Experiments in segmenting texton patterns using localized spatial filters," *Pattern Recognit.*, vol. 22, no. 6, pp. 707–717, 1989.

[36] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[37] Y. El-Shamayleh, R. D. Kumbhani, N. T. Dhruv, and J. A. Movshon, "Visual response properties of v1 neurons projecting to v2 in macaque," *J. Neurosci.*, vol. 33, no. 42, pp. 16594–16605, 2013.

[38] J. Hegdé and V. E. Van Essen, "Selectivity for complex shapes in primate visual area v2," *J Neurosci*, vol. 20, no. 5, pp. 61–66, 2000.
[39] A. Anzai, X. Peng, and D. C. Van Essen, "Neurons in monkey visual area v2 encode combinations of orientations," *Nature Neurosci.*, vol. 10, no. 10, pp. 1313–1321, 2007.
[40] M. Ito and H. Komatsu, "Representation of angles embedded within contour stimuli in area v2 of macaque monkeys," *J. Neurosci.*, vol. 24, no. 13, pp. 3313–3324, 2004.
[41] B. D. B. Willmore, R. J. Prenger, and J. L. Gallant, "Neural representation of natural images in visual area v2," *J. Neurosci.*, vol. 30, no. 6, pp. 2102–2114, 2010.
[42] A. M. Schmid, K. P. Purpura, I. E. Ohiorhenuan, F. Mechler, and J. D. Victor, "Subpopulations of neurons in visual area v2 perform differentiation and integration operations in space and time," *Frontiers Syst. Neurosci.*, vol. 3, p. 15, Nov. 2009.
[43] J. Freeman and E. P. Simoncelli, "Metamers of the ventral stream," *Nature Neurosci.*, vol. 14, no. 9, pp. 1195–1201, 2011.
[44] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.
[45] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, vol. 2. Nov. 2003, pp. 1398–1402.
[46] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.
[47] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312–322, Sep. 2004.
[48] P. V. Vu, C. T. Vu, and D. M. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2505–2508.
[49] R. T. Born and D. C. Bradley, "Structure and function of visual area MT," *Annu. Rev. Neurosci.*, vol. 28, pp. 157–189, Jul. 2005.
[50] E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area MT," *Vis. Res.*, vol. 38, no. 5, pp. 743–761, 1998.
[51] J. A. Perrone, "A visual motion sensor based on the properties of V1 and MT neurons," *Vis. Res.*, vol. 44, no. 15, pp. 1733–1755, 2004.

Anish Mittal received the B.Tech. degree in electrical engineering from IIT Roorkee, in 2009, and the M.S. and Ph.D. degrees in electrical and computer engineering from The University of Texas at Austin, in 2011 and 2013, respectively. He is currently a Senior Engineer and Researcher with Reality Capture and Processing team, HERE maps, Berkeley. His research interests include applied multidimensional signal processing, statistical modeling, and machine learning.

Michele A. Saad received the B.E. degree in computer and communications engineering from the American University of Beirut, Lebanon, in 2007, and the M.S. and Ph.D. degrees in electrical and computer engineering from The University of Texas at Austin, in 2009 and 2013, respectively. She is currently a Senior Engineer and Researcher of Perceptual Image and Video Engineering with Intel. Her research interests include statistical modeling of images and videos, motion perception, design of perceptual image and video quality assessment algorithms, and statistical modeling, data mining, and machine learning. She is the Co-Chair of the Video/Image Models for Consumer Content Evaluation project at the Video Quality Experts Group.

Alan C. Bovik (F'96) is currently the Cockrell Family Endowed Regents Chair in Engineering at The University of Texas at Austin, where he is the Director of the Laboratory for Image and Video Engineering with the Department of Electrical and Computer Engineering and the Institute for Neuroscience. He has authored over 750 technical articles in these areas and holds several U.S. patents. His publications have been cited more than 45 000 times in the literature, his current H-index is about 75, and he is listed as a Highly-Cited Researcher by Thompson Reuters. His research interests include image and video processing, digital television and digital cinema, computational vision, and visual perception. His several books include the companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, 2009). He received a Primetime Emmy Award for Outstanding Achievement in Engineering Development from the Television Academy in 2015, for his work on the development of video quality prediction models which have become standard tools in broadcast and post-production houses throughout the television industry. He has also received a number of major awards from the IEEE Signal Processing Society, including: the Society Award (2013); the Technical Achievement Award (2005); the best paper award (2009); the Signal Processing Magazine Best Paper Award (2013); the Education Award (2007); the Meritorious Service Award (1998), and (co-author) the Young Author Best Paper Award (2013). He also was named recipient of the Honorary Member Award of the Society for Imaging Science and Technology for 2013, received the SPIE Technology Achievement Award for 2012, and was the IS&T/SPIE Imaging Scientist of the Year for 2011. He is also a recipient of the Joe J. King Professional Engineering Achievement Award (2015) and the Hocott Award for Distinguished Engineering Research (2008), both from the Cockrell School of Engineering at The University of Texas at Austin, and the Distinguished Alumni Award from the University of Illinois at Champaign–Urbana (2008). He is a fellow of the Optical Society of America and the Society of Photo-Optical and Instrumentation Engineers, and a member of the Television Academy, the National Academy of Television Arts and Sciences, and the Royal Society of Photography. He also co-founded and was the longest-serving Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING (1996-2002), and created and served as the first General Chair of the IEEE International Conference on Image Processing, held in Austin, TX, in 1994. His many other professional society activities include: Board of Governors, IEEE Signal Processing Society, 1996-1998; Editorial Board, THE PROCEEDINGS OF THE IEEE, 1998-2004; and Series Editor for *Image, Video, and Multimedia Processing*, Morgan and Claypool Publishing Company, 2003-present. His was also the General Chair of the 2014 Texas Wireless Symposium, held in Austin in 2014. He is a registered Professional Engineer in the state of Texas and a Frequent Consultant to legal, industrial, and academic institutions.