Video Quality Assessment by Reduced Reference Spatio-Temporal Entropic Differencing

Rajiv Soundararajan and Alan C. Bovik, *Fellow, IEEE*

***Abstract*—We present a family of reduced reference video quality assessment (QA) models that utilize spatial and temporal entropic differences. We adopt a hybrid approach of combining statistical models and perceptual principles to design QA algorithms. A Gaussian scale mixture model for the wavelet coefficients of frames and frame differences is used to measure the amount of spatial and temporal information differences between the reference and distorted videos, respectively. The spatial and temporal information differences are combined to obtain the spatio-temporal-reduced reference entropic differences. The algorithms are flexible in terms of the amount of side information required from the reference that can range between a single scalar per frame and the entire reference information. The spatio-temporal entropic differences are shown to correlate quite well with human judgments of quality, as demonstrated by experiments on the LIVE video quality assessment database.**

***Index Terms*—Entropy, human visual system, motion information, natural video statistics, reduced reference video quality assessment.**

I. Introduction

THE IMPACT of digital video on today's life is pervasive with applications ranging from video streaming over mobile devices and computers to video conferencing, surveillance, and digital cinema. This renders the question of user experience through quality monitoring and control extremely important. Image and video quality assessment (QA) concerns the problem of quantifying this user experience through objective algorithms that implement quality indices designed to correlate well with subjective judgments of quality. This paper deals with reduced reference (RR) QA, where only partial information from the reference can be made available in addition to the distorted video to conduct quality evaluation.

The problem of RR QA becomes particularly relevant in the context of video owing to the large amount of data involved. RR video quality assessment (VQA) is a challenging research problem for two reasons: the RR constraint in the problem and the multidimensional (spatial and temporal) structure of the signal. Elaborating on the second aspect, VQA is significantly harder than image quality assessment (IQA) owing to the addition of a new dimension to the problem, namely, the temporal dimension. A key attribute of the most successful VQA algorithms is their ability to capture both temporal and spatial distortions. Blocking artifacts, ringing effects, and blur are the examples of spatial distortions, while jerkiness, ghosting, and mosquito effects are the examples of temporal distortions. Reference [1] contains detailed descriptions of various spatial and temporal distortions that can afflict a video signal.

The spatial aspect of the problem has received a lot of attention over the years, and significant progress has been made owing to the availability of sophisticated models of the human visual system (HVS) and of natural scene statistics. However, the successful application of either HVS-based approaches or statistical approaches to capture temporal distortions has been limited. While there do exist advanced models of motion perception, their applicability to VQA has not yet been completely realized. References [1] and [2] represent the examples of an effort toward this direction. On the other hand, statistical models of motion are almost nonexistent in the literature and none has been found to be statistically regular over natural videos. Motion and other temporal changes in a video may be analyzed through optical flow vectors or by temporal filtering and decomposition. Reference [3] provides a statistical model of optical flow vectors. However, the regularity of this model over all kinds of videos, including those that contain egomotion or significant movements over large areas of the frame, appears to be difficult to verify. There is also little prior work on characterizing the statistics of temporally filtered natural videos. In this paper, we build a statistical model of multiscale multiorientation wavelet decompositions of frame differences. This represents one of the first attempts at trying to obtain a statistical model of temporal changes in natural videos and to use them for VQA.

The main contributions of this paper are as follows.

1) We present a natural video statistical model for the wavelet coefficients of frame differences between the adjacent frames in a video sequence. These coefficients possess a heavy tailed distribution and are well modeled by a Gaussian scale mixture (GSM) distribution.
2) The GSM model for the wavelet coefficients of frame differences is used to compute RR entropic differences

Manuscript received December 22, 2011; revised March 18, 2012, May 18, 2012, June 19, 2012, and July 31, 2012; accepted August 2, 2012. Date of publication August 22, 2012; date of current version April 1, 2013. This work was supported in part by the National Science Foundation under Grant CCF-0916713, the Air Force Office of Scientific Research (AFOSR) under Grant FA95500910063, and Intel Corporation and Cisco Systems, Inc., under the Video Aware Wireless Networks (VAWN) Program. This paper was recommended by Associate Editor L. Zhang.

R. Soundararajan was with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA. He is now with Qualcomm Research India, Bangalore 560066, India (e-mail: rajivs@utexas.edu).

A. C. Bovik is with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA (e-mail: 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/TCSVT.2012.2214933

(RRED) between the reference and the distorted videos, leading to temporal RRED (TRRED) indices. These indices are designed using a hybrid approach of statistical models and perceptual principles. The TRRED indices seek to measure the amount of motion information difference that occurs between the reference and distorted videos. We hypothesize that this information difference captures temporal distortions that can be perceived by humans.

3) The TRRED indices, in conjunction with our previously developed spatial RRED (SRRED) indices evaluated by applying the RRED index [4] on every frame of the video, yield the spatio-temporal RRED (STRRED) index, which performs very well on the LIVE video quality assessment database in terms of correlation with human judgments of quality.
4) A family of algorithms are developed that vary in the amount of information required from the reference for quality computation. In particular, single-number algorithms or indices that require just a single number from the reference or distorted video per frame are developed, which are shown to correlate well with human judgments of quality. The amount of information can range up to almost full reference.

The SRRED, TRRED, and STRRED indices possess all the favorable properties of the RRED indices in [4]. In particular, they allow for bidirectional computation of quality, need not be trained on databases of human opinion scores, and are applicable to scalable VQA. Depending on the desired application and problem constraints, this framework of algorithms also allows users to pick any algorithm from the class of algorithms presented which meets the performance requirements and data rates required for quality computation.

We now present a brief overview of prior work on video QA. A detailed subjective evaluation of popular full reference VQA algorithms can be found in [5]. A classification and subjective evaluation of full reference and RR VQA algorithms is also contained in [6], while [7] provides a tutorial on VQA. Any image QA algorithm can be trivially extended to videos by applying the algorithm on every frame of the video and by calculating the average score. Thus, successful IQA algorithms, such as multiscale structural similarity index (SSIM) [8], Sarnoff just noticeable differences matrix [9], and visual signal-to-noise ratio [10], naturally lead to very simple VQA algorithms. However, the performance of these algorithms is limited owing to their failure to capture temporal distortions in the video. Various researchers have developed QA algorithms that attempt to capture both spatial and temporal distortions. Reference [2] computes a speed weighted SSIM index by associating weights with different locations depending on the amount of motion information. The MOVIE index [1] uses the idea of motion tuning or the varied sensitivity of the human to motion direction to measure distortions. Other FR VQA algorithms account for temporal distortions from temporal filter responses [11], temporal decorrelation [12], evaluating quality over temporal trajectories [13] and studying the temporal evolution of spatial distortions [14].

The video quality metric (VQM) introduced by NTIA [15] is an RR VQA algorithm that requires reference data of around 4% of the size of the uncompressed video sequence. The algorithm is based on computing losses in the spatial gradients of the luminance components and features based on the product of luminance contrast and motion, and on measuring color impairments. Reference [16] is another RR VQA algorithm that calculates the weighted norm of the error in wavelet coefficients at selected locations to reduce the data required for quality computation. Other recent RR VQA algorithms include [17] and [18]. While the former is based on a discriminative analysis of harmonic strength, the latter is an extension of [2] to the RR setting. The STRRED indices introduced in this paper offer significant improvements in performance and allow for excellent performance at low data rates of reference side information.

The remainder of this paper is organized as follows. We present a statistical model for the wavelet coefficients of frame differences in Section II, followed by the system model on which the QA algorithms are based in Section III. We describe the STRRED indices in Section IV, provide a perceptual interpretation of the algorithms in Section V, and discuss the performance of these indices in Section VI. We conclude this paper in Section VII.

II. Statistical Model of Frame Differences

We describe a statistical model of the wavelet coefficients (WC) of the frame differences between the adjacent frames in a video sequence. The wavelet coefficients are obtained by a steerable pyramid decomposition at multiple scales and orientations [19]. Let all the wavelet coefficients of the frame differences in a subband be partitioned into nonoverlapping blocks of size $\sqrt{N} \times \sqrt{N}$. Let the blocks in subband k, $k \in \{1, 2, \ldots, K\}$, be indexed by $m \in \{1, 2, \ldots, M_k\}$. Now, consider the wavelet coefficients of the frame differences between frames f and $f+1$ belonging to block m of subband k. Let $\bar{D}_{mkf}$ denote a vector of wavelet coefficients in block m of subband k and frame f.

We model the block $\bar{D}_{mkf}$ as a GSM distributed vector with a continuous scale. Specifically, $\bar{D}_{mkf}$ is distributed as follows:

$$\bar{D}_{mkf} = T_{mkf} \bar{V}_{mkf} \tag{1}$$

where T_{mkf} is independent of $\bar{V}_{kf}$ with $\bar{V}_{kf} \sim \mathcal{N}(0, \mathbf{K}_{V_{kf}})$. Note that we model $\bar{V}_{mkf}$ in every block of subband k and frame f to have the same covariance matrix $\mathbf{K}_{kf}$, with the premultiplier random variable T_{mkf} modulating the covariance matrix for different blocks. We also assume that T_{mkf} and $\bar{V}_{kf}$ are independent across all the indices m, k, and f describing them.

Note that similar models describe the wavelet coefficients of frames in [20]. The GSM model implies that the divisively normalized coefficients (i.e., coefficients obtained by dividing $\bar{D}_{mkf}$ by T_{mkf}) follow a Gaussian distribution. This implication is clear since $\bar{D}_{mkf}/T_{mkf} = \bar{V}_{mkf}$, which is distributed as $\mathcal{N}(0, \mathbf{K}_{V_{kf}})$. We verify this empirically by obtaining the maximum likelihood (ML) estimator of T_{mkf}, dividing the wavelet coefficient blocks by these, and analyzing the distributions of

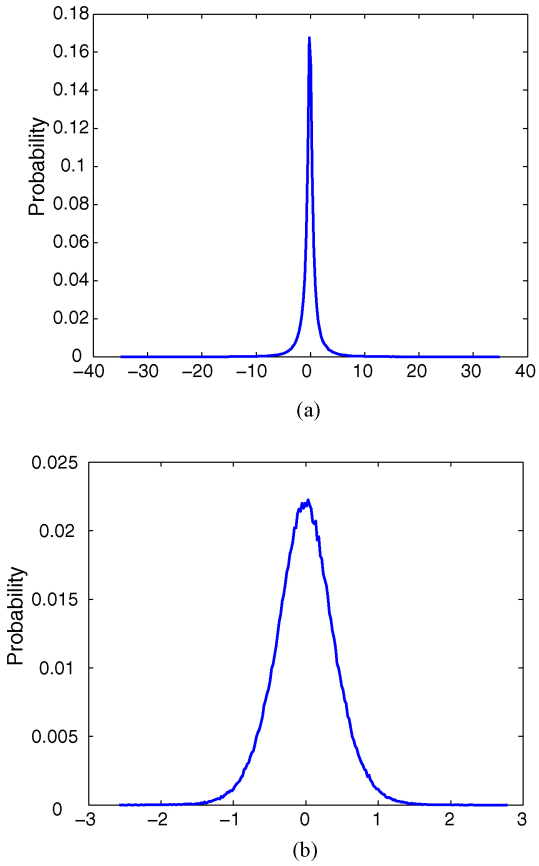

Fig. 1. Distributions of wavelet coefficients of frame differences in a subband of reference videos. (a) *Sunflower*. (b) *Mobile Calendar*.

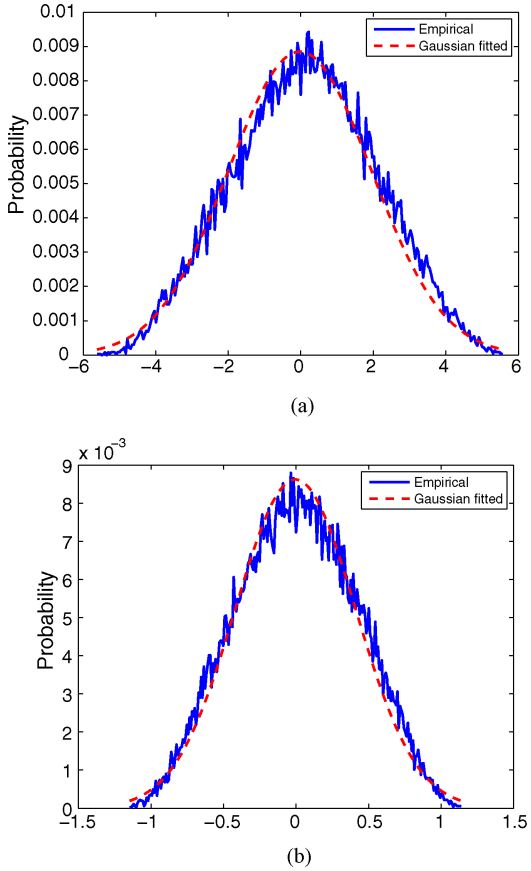

Fig. 2. Empirical and Gaussian fitted statistics of divisively normalized wavelet coefficients (in a subband) of frame differences between 10th and 11th frames of *Sunflower* and *Mobile Calendar* video sequences in the LIVE video quality assessment database. ΔH denotes the relative entropy between the empirical distribution and the Gaussian fitted distribution, while H denotes the entropy of the empirical distribution of the normalized coefficients. (a) *Sunflower*, $\frac{\Delta H}{H} = 0.0030$. (b) *Mobile Calendar*, $\frac{\Delta H}{H} = 0.0029$.

the normalized coefficients. The ML estimator of T_{mkf} is given by [20]

$$\hat{T}_{mkf} = \underset{T_{mkf}}{\operatorname{argmax}}\, p(\bar{D}_{mkf}|T_{mkf}) = \sqrt{\frac{\bar{D}_{mkf}^T \mathbf{K}_{V_{kf}}^{-1} \bar{D}_{mkf}}{N}}.$$

In Fig. 1, we show that the empirical histograms of the WC of the frame differences are non Gaussian and heavy tailed. In Fig. 2, we show that the corresponding normalized coefficients can be modeled well by a Gaussian distribution. The ratio of the relative entropy between the empirical distribution of the normalized coefficients and the Gaussian fitted distribution (denoted by ΔH) is calculated and shown to be a very small fraction of the entropy of the corresponding empirical distributions (denoted by H). Further, in Table I, we show that the ratio of relative entropy to the entropy of the empirical distribution ($\Delta H/H$) computed for each frame and averaged over all frames in every reference video sequence on the LIVE VQA database [5] is small. In the remainder of this paper, we refer to the local variances of the wavelet coefficients of the frame differences as local temporal variances.

In Figs. 3 and 4, we show the exemplar corresponding plots (of the empirical histograms and the normalized coefficients) for videos distorted by H.264 compression. It is evident that the normalized coefficients of *Sunflower* appear to be Gaussian distributed, while those of *Mobile Calendar* are not Gaussian distributed. While we do not explicitly exploit the Gaussianity or non-Gaussianity of the normalized coefficients of the distorted videos in our algorithms, we use a GSM model for the distorted video's temporal coefficients, as was done in [4] for the spatial coefficients of the distorted images. By fitting a GSM model for the distorted video, we compute the natural approximation of the distorted video. If the normalized coefficients are, indeed, Gaussian distributed, then it means that the natural approximation is close to the empirical distributions of the distorted coefficients. The entropy difference between the reference and the distorted videos can then be interpreted as a distance between the reference video and the natural approximation of the distorted video. Our model is based on the hypothesis that such a distance is meaningful for perceptual quality, and this is validated by the good performance of our algorithms on the LIVE VQA database. This also avoids the need for learning distortion-specific statistical models.

III. System Model

Let $\bar{C}_{mkfr}$ and $\bar{C}_{mkfd}$ denote a vector of wavelet coefficients obtained through a steerable pyramid decomposition [19] in block m, subband k, and frame f of the reference and distorted videos, respectively. On account of the GSM model for wavelet

TABLE I
MEAN $\Delta H/H$ FOR REFERENCE VIDEOS ON THE LIVE VQA DATABASE

Sequence	1	2	3	4	5	6	7	8	9	10
Mean $\Delta H/H$	0.0030	0.0019	0.0037	0.0024	0.0017	0.0025	0.0022	0.0024	0.0034	0.0026

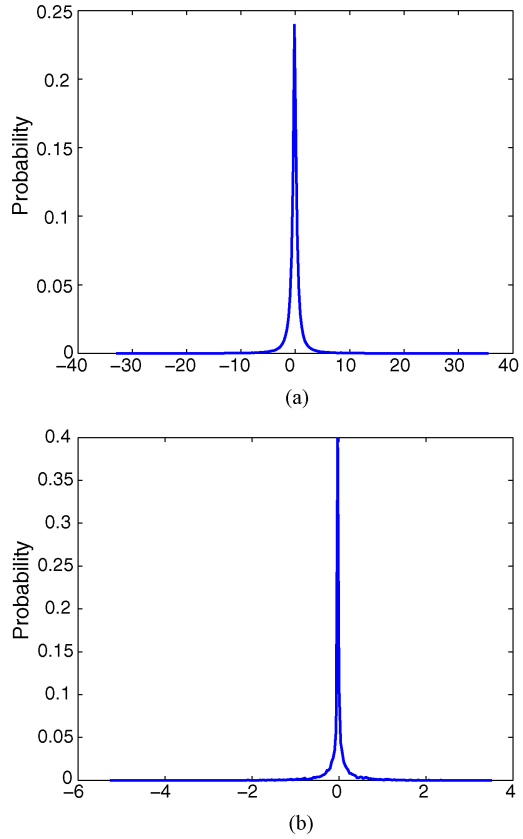

Fig. 3. Distributions of the wavelet coefficients of the frame differences in a subband of H.264 compressed videos. (a) *Sunflower*. (b) *Mobile Calendar*.

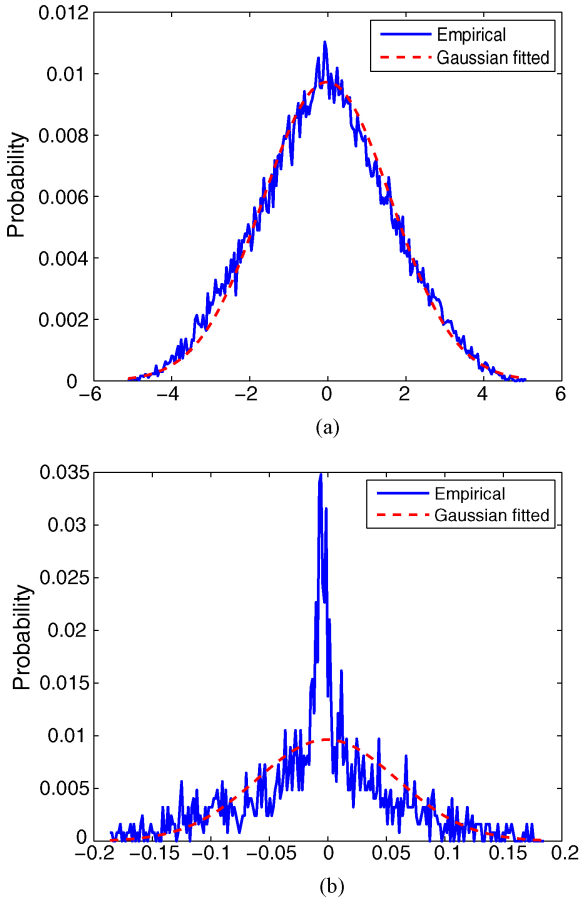

Fig. 4. Empirical and Gaussian fitted statistics of divisively normalized wavelet coefficients (in a subband) of frame differences between 10th and 11th frames of (a) H.264 compressed *Sunflower* and (b) H.264 compressed *Mobile Calendar* video sequences in the LIVE video quality assessment database.

coefficients, we have

$$\bar{C}_{mkfr} = S_{mkfr}\bar{U}_{mkfr} \qquad \bar{C}_{mkfd} = S_{mkfd}\bar{U}_{mkfd}$$

where S_{mkfr} is independent of $\bar{U}_{mkfr}$, S_{mkfd} is independent of $\bar{U}_{mkfd}$, $\bar{U}_{mkfr} \sim \mathcal{N}(0.\mathbf{K}_{U_{kfr}})$, $\bar{U}_{mkfd} \sim \mathcal{N}(0.\mathbf{K}_{U_{kfd}})$, and S_{mkfr} and S_{mkfd} are nonnegative random variables. We assume that all the blocks are independent of each other in order to simplify the index. Interblock interactions are not accounted for in this index, although it is likely possible to obtain better indices by exploiting such interactions as well, given that accurate enough models could be found to apply. Note that as in [4], we use a natural image approximation for the WC in each frame of the distorted video as well and measure quality using this natural video approximation.

Let $\bar{D}_{mkfr}$ and $\bar{D}_{mkfd}$ be a vector of wavelet coefficients of frame differences between frames f and $f+1$ in block m and subband k belonging to the reference and distorted videos, respectively. Using the model in (1) for the WC of frame differences, we have

$$\bar{D}_{mkfr} = T_{mkfr}\bar{V}_{mkfr} \qquad \bar{D}_{mkfd} = T_{mkfd}\bar{V}_{mkfd}$$

where T_{mkfr} is independent of $\bar{V}_{kfr}$, T_{mkfd} is independent of $\bar{V}_{kfd}$, $\bar{V}_{kfr} \sim \mathcal{N}(0.\mathbf{K}_{V_{kfr}})$, $\bar{V}_{kfd} \sim \mathcal{N}(0.\mathbf{K}_{V_{kfd}})$, and T_{mkfr} and T_{mkfd} are nonnegative random variables. Again, as before, all the blocks are assumed to be independent of each other.

We allow the wavelet coefficients in each block of the reference and distorted video frames, as well as the wavelet coefficients of the frame differences corresponding to each block of the reference and distorted frames to pass through a Gaussian channel in order to model imperfections in the visual perception of these coefficients in the human visual system, e.g., neural noise [21]. These models are shown in Fig. 5 and are expressed as follows:

$$\bar{C}'_{mkfr} = \bar{C}_{mkfr} + \bar{W}_{mkfr} \qquad \bar{C}'_{mkfd} = \bar{C}_{mkfd} + \bar{W}_{mkfd}$$

where $\bar{W}_{mkfr} \sim \mathcal{N}(0, \sigma_W^2\mathbf{I}_N)$ and $\bar{W}_{mkfd} \sim \mathcal{N}(0, \sigma_W^2\mathbf{I}_N)$.

Similarly, we have

$$\bar{D}'_{mkfr} = \bar{D}_{mkfr} + \bar{Z}_{mkfr} \qquad \bar{D}'_{mkfd} = \bar{D}_{mkfd} + \bar{Z}_{mkfd}$$

where $\bar{Z}_{mkfr} \sim \mathcal{N}(0, \sigma_Z^2\mathbf{I}_N)$ and $\bar{Z}_{mkfd} \sim \mathcal{N}(0, \sigma_Z^2\mathbf{I}_N)$.

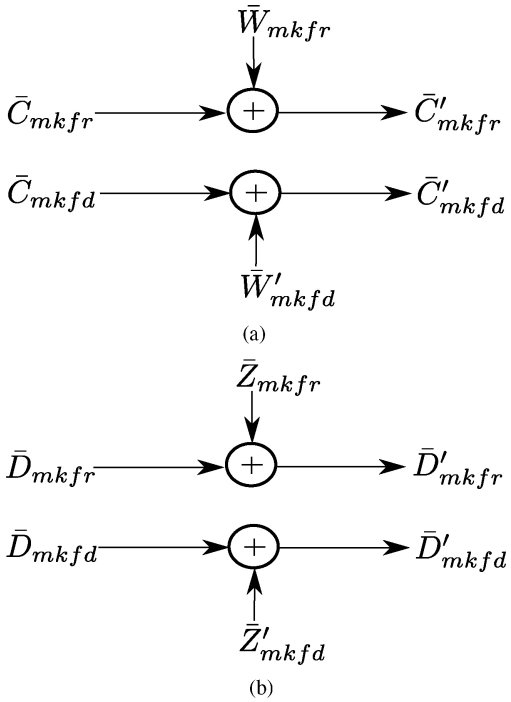

Fig. 5. System model for computation of SRRED and TRRED indices. (a) Wavelet coefficients of frames. (b) Wavelet coefficients of frame differences.

The RRED indices essentially compute the differences of scaled entropies of the neural noisy wavelet coefficients of frames or neural noisy wavelet coefficients of the frame differences. These are described in detail in the following section.

IV. STRRED Indices

We first describe the design of the SRRED and TRRED indices and then show how these may be combined to yield STRRED indices. Both SRRED and TRRED indices are based on conditional entropic differences using the GSM model for wavelet coefficients. The entropy is conditioned on the realization of the premultiplier random variable of the GSM model being its ML estimate. See [4] for a discussion on the perceptual relevance of conditional entropies. The conditional entropy calculations here require the GSM model and it does not appear to be possible to evaluate these using just empirical distributions. We also do not use simple local entropic differences without the conditioning since such an algorithm would involve unreliable estimation of joint densities (since only one sample is available for each block) in the entropy computations.

A. SRRED

The SRRED indices are computed in a manner similar to [4]. We evaluate the entropies of $\bar{C}'_{mkfr}$ and $\bar{C}'_{mkfd}$ conditioned on the ML estimates of S_{mkfr} and S_{mkfd}, respectively. Let s_{mkfr} and s_{mkfd} be the ML estimates of S_{mkfr} and S_{mkfd} given the corresponding frames in the reference and distorted videos, respectively. The entropies of $\bar{C}'_{mkfr}$ and $\bar{C}'_{mkfd}$ conditioned on $S_{mkfr} = s_{mkfr}$ and $S_{mkfd} = s_{mkfd}$ are given by

$$h(\bar{C}'_{mkfr}|S_{mkfr} = s_{mkfr}) = \frac{1}{2}\log\left[(2\pi e)^N |s^2_{mkfr}\mathbf{K}_{U_{kfr}} + \sigma^2_W \mathbf{I}_N|\right]$$

$$h(\bar{C}'_{mkfd}|S_{mkfd} = s_{mkfd}) = \frac{1}{2}\log\left[(2\pi e)^N |s^2_{mkfd}\mathbf{K}_{U_{kfd}} + \sigma^2_W \mathbf{I}_N|\right].$$

Let the scaling factors be defined by

$$\gamma_{mkfr} = \log(1 + s^2_{mkfr}) \qquad \gamma_{mkfd} = \log(1 + s^2_{mkfd}). \tag{2}$$

The scalars defined in (2) are exactly the same as used in [4], and enhance the local nature of the algorithm, allow for variable weighting of the amount of visual information in different regions of each frame, and embed numerical stability in the algorithm for small values of the neural noise variance and the local spatial variance estimate.

Let

$$\alpha_{mkfr} = \gamma_{mkfr} h(\bar{C}'_{mkfr}|S_{mkfr} = s_{mkfr})$$
$$\alpha_{mkfd} = \gamma_{mkfd} h(\bar{C}'_{mkfd}|S_{mkfr} = s_{mkfd}).$$

The SRRED index in subband k obtained by using M_k scalars (one for each block) from the reference and distorted videos is given by

$$\text{SRRED}^{M_k}_k = \frac{1}{FM_k}\sum_{f=1}^{F}\sum_{m=1}^{M_k}|\alpha_{mkfr} - \alpha_{mkfd}|.$$

The number of scalars required in subband k can be reduced by summing the scaled entropy terms over small patches and sending these partial sums. The number of scalars required is equal to the number of patches. As an extreme case, all the entropy terms can be summed up and sent as a single number from the reference. The SRRED index in subband k using one scalar is given by

$$\text{SRRED}^1_k = \frac{1}{FM_k}\sum_{f=1}^{F}\left|\sum_{m=1}^{M_k}\alpha_{mkfr} - \sum_{m=1}^{M_k}\alpha_{mkfd}\right|.$$

B. TRRED

The TRRED indices are obtained by computing the differences of scaled conditional entropies of the wavelet coefficient differences. Denote t_{mkfr} and t_{mkfd} as the ML estimates of T_{mkfr} and T_{mkfd} given the corresponding adjacent frames in the reference and distorted videos, respectively. The entropies of $\bar{D}'_{mkfr}$ and $\bar{D}'_{mkfd}$ conditioned on $T_{mkfr} = t_{mkfr}$ and $T_{mkfd} = t_{mkfd}$ are given by

$$h(\bar{D}'_{mkfr}|T_{mkfr} = t_{mkfr}) = \frac{1}{2}\log\left[(2\pi e)^N |t^2_{mkfr}\mathbf{K}_{V_{kfr}} + \sigma^2_Z \mathbf{I}_N|\right]$$

$$h(\bar{D}'_{mkfd}|T_{mkfd} = t_{mkfd}) = \frac{1}{2}\log\left[(2\pi e)^N |t^2_{mkfd}\mathbf{K}_{V_{kfd}} + \sigma^2_Z \mathbf{I}_N|\right].$$

Now, define the scaling factors

$$\delta_{mkfr} = \log(1 + s^2_{mkfr})\log(1 + t^2_{mkfr})$$
$$\delta_{mkfd} = \log(1 + s^2_{mkfd})\log(1 + t^2_{mkfd}) \tag{3}$$

which attach different importance to the amount of visual information at different locations and depend on both local temporal and spatial variance parameters in contrast to the scaling factors defined in (2), which only depend on the spatial variance parameters. The scalars used in (3) are a product of two components. The first component, which is a function of the local variance of the spatial decomposition of the frame, can be interpreted as follows. References [22] and [23] suggest that at slow speeds, lowering the contrast lowers the perceived speed. Under the assumption that the speeds due to frames sampled at practical temporal sampling rates (between 25 and 50 f/s) can be considered slow speeds, we believe that the factor $\log(1+s^2_{mkfr})$ (or $\log(1+s^2_{mkfd})$), which is an increasing function of the local spatial variance, lowers the perceived speed as the contrast reduces. The local spatial variance s^2_{mkfr} (or s^2_{mkfd}) measures the spatial luminance contrast in different regions of the frame. The other component of the scaling factor, $\log(1+t^2_{mkfr})$ (or $\log(1+t^2_{mkfd})$), which depends on the local temporal variance, has an effect similar to the effect of local spatial variances for SRRED indices, where it embeds local nature to the algorithm and allows weighting of the temporal information from different locations according to the amount of local temporal variance.

Let

$$\beta_{mkfr} = \delta_{mkfr} h(\bar{D}'_{mkfr} | T_{mkfr} = t_{mkfr})$$
$$\beta_{mkfd} = \delta_{mkfd} h(\bar{D}'_{mkfd} | T_{mkfr} = t_{mkfd}).$$

The TRRED index in subband k obtained by using M_k scalars from the reference and distorted video difference frames is given by

$$\text{TRRED}_k^{M_k} = \frac{1}{FM_k} \sum_{f=1}^{F} \sum_{m=1}^{M_k} |\beta_{mkfr} - \beta_{mkfd}|.$$

Similar to the SRRED indices, the amount of information required in subband k of frame f for the TRRED index can be reduced by summing the entropy terms over small patches. In the limiting case, we obtain single-number algorithms, where we require one number per frame for the evaluation of the TRRED index. This may be represented as follows:

$$\text{TRRED}_k^{1} = \frac{1}{FM_k} \sum_{f=1}^{F} \left| \sum_{m=1}^{M_k} \beta_{mkfr} - \sum_{m=1}^{M_k} \beta_{mkfd} \right|.$$

C. STRRED

The STRRED indices combine the SRRED and the TRRED indices. Note that the SRRED and TRRED indices operate individually on data obtained by separate processing of the spatial and temporal frequency components. This matches well-accepted models of separable spatial and temporal frequency tuning of area V1 neurons [24], [25]. According to this model, the response of area V1 neurons to temporal frequencies is not affected by the spatial frequency of the stimulus and vice versa. A product form can thus be used to represent separable spatial and temporal frequency responses of the area V1 cortical neurons. We model the separable processing of area V1 neurons as opposed to [1], which models the behavior of the neurons in area middle temporal (MT). The area MT neurons are tuned to velocity (both direction and speed), and their selectivity toward motion direction and speed inhibits the spatio-temporal frequency separability. By matching the separable part of the cortical processing, we are able to capture pure temporal distortions that are often flickery without computing motion vectors. This is perhaps, in particular, relevant and important in the context of QA.

Interestingly, while the SRRED indices are obtained using only spatial frequency information, the TRRED indices are obtained using spatial and temporal information (the spatial information is used to weigh the temporal information). As a result, only the TRRED indices are influenced by temporal distortions, while both SRRED and TRRED indices are affected by spatial distortions. The computation of the quality index from the spatial and temporal information concerns the processing that occurs in the later stages of human visual processing, where there is evidence of interactions between the two [24]. While we are inspired by these observations, and have used them in constructing our RR VQA models, we do not claim to replicate specific cortical processing modules.

The STRRED index is obtained as a product of the SRRED and TRRED indices and is expressed as follows:

$$\text{STRRED}_k = \text{SRRED}_k \text{TRRED}_k.$$

Although the TRRED index, indeed, does involve the local spatial variances through the scaling factor, it still only computes temporal entropic differences while only the SRRED index computes spatial entropic differences. Thus, it is the combination of both SRRED and TRRED indices that renders the STRRED index an effective VQA algorithm, especially in the regime of low reference information as shown in Section VI.

D. Estimation of Parameters

The ML estimates of the local spatial and temporal variances as well as the covariance matrices in a given subband are given below. Derivations of these estimates can be found in [20] and [21]. The estimates of the spatial and temporal covariance matrices of subband k in frame f of the reference and distorted videos are given by

$$\hat{\mathbf{K}}_{U_{kfr}} = \sum_{m=1}^{M_k} \frac{\bar{C}_{mkfr} \bar{C}^T_{mkfr}}{M_k} \quad \hat{\mathbf{K}}_{V_{kfr}} = \sum_{m=1}^{M_k} \frac{\bar{D}_{mkfr} \bar{D}^T_{mkfr}}{M_k}$$
$$\hat{\mathbf{K}}_{U_{kfd}} = \sum_{m=1}^{M_k} \frac{\bar{C}_{mkfd} \bar{C}^T_{mkfd}}{M_k} \quad \hat{\mathbf{K}}_{V_{kfd}} = \sum_{m=1}^{M_k} \frac{\bar{D}_{mkfd} \bar{D}^T_{mkfd}}{M_k}.$$

Similarly, the ML estimates of the local spatial and temporal

variances of the reference and distorted videos are given by

$$\hat{s}^2_{mkfr} = \frac{\bar{C}^T_{mkfr}\mathbf{K}^{-1}_{U_{kfr}}\bar{C}_{mkfr}}{N}$$
$$\hat{t}^2_{mkfr} = \frac{\bar{D}^T_{mkfr}\mathbf{K}^{-1}_{V_{kfr}}\bar{D}_{mkfr}}{N}$$
$$\hat{s}^2_{mkfd} = \frac{\bar{C}^T_{mkfd}\mathbf{K}^{-1}_{U_{kfd}}\bar{C}_{mkfd}}{N}$$
$$\hat{t}^2_{mkfd} = \frac{\bar{D}^T_{mkfd}\mathbf{K}^{-1}_{V_{kfd}}\bar{D}_{mkfd}}{N}.$$

V. Perceptual Properties of STRRED Indices

We now discuss the perceptual properties that the STRRED indices rely on. See [4] for more discussion on perceptual aspects of the algorithm that relate to the STRRED indices developed here.

A. Spatial and Temporal Multiscale Multiorientation Decomposition

The STRRED indices involve separate spatial and temporal decompositions of the given video motivated by the evidence of mostly separable processing of the spatial and temporal data in the visual cortex. In particular, the SRRED indices are computed using spatial multiscale multiorientation decompositions of each frame in the video sequence, while the TRRED indices are computed using a spatial multiscale multiorientation decomposition of the frame differences. Frame differences capture the temporal information in the video, and we further subject it to a multiscale multiorientation decomposition before computing the index. We discuss the perceptual implications of each of these aspects in the following.

The SRRED indices are computed according to [4] and therefore involve a multiscale multiorientation decomposition of the given image before the index is computed. There is ample evidence in the visual science literature that suggests that similar signal processing occurs in the early stages of visual processing. We refer the reader to [1] and [4] for a detailed account of such signal processing and a comparison of various filters used in this process. Here, we simply mention that steerable pyramids are used for the multiscale multiorientation spatial wavelet decomposition as opposed to the spatio-temporal Gabor filters used in [1].

We capture temporal information present in the video signal through frame differences and subject them to a multiscale multiorientation wavelet decomposition using steerable pyramids. High-magnitude wavelet coefficients of frame differences may be interpreted as moving edges, and we substantiate this conclusion through the experimental results shown in Table II. The objective of the experiment is to find out the relationship between the local variances of the wavelet coefficients of the frame differences (denoted by T and referred to as local temporal variances), the local variances of wavelet coefficients of frames (denoted by S and referred to as local spatial variances), and the optical flow evaluated at each block corresponding to S and T using the algorithm in [26] (denoted by OF). Note that $\mathbb{E}[T] \sim 1$, as indicated in the estimation procedure in the previous section. We show through Table II that $\mathbb{E}[S|T > 1] \geq \mathbb{E}[S]$ and $\mathbb{E}[OF|T > 1] \geq \mathbb{E}[OF]$ for the ten diverse reference videos on the LIVE video quality assessment database. In Table II, the empirical estimates of $\mathbb{E}[S]$ and $\mathbb{E}[S|T > 1]$ are obtained for each frame after computing the ML estimate of S and T at every block, and the estimates of the expectations in each frame are again averaged across frames for every reference video. The flow estimates are obtained using the optical flow algorithm [26] and the empirical expectations are evaluated in a similar manner. Similar behavior is obtained even when the expectations are conditioned on the event $T > c$ with $c > 1$. These results suggest that given that T is large, S and OF are also large. S is large at high frequency locations corresponding to edges, while OF is large in locations where there is motion. Thus T being high implies that S and OF are high. Therefore, locations where T is high may be interpreted as moving edges. Note that in order to interpret the meaning of large temporal variance, we study the behavior of spatial variance and optical flow when the temporal variance is high. We do not make any claims about the temporal variance when either the spatial variance or the optical flow is large.

B. Effect of Motion on Spatio-Temporal Information

Motion tends to have the effect of shifting the frequency response curves down the scale of the wavelet decomposition. According to [27], "motion does not diminish the visual passband, but instead slides the spatial frequency window down the spatial frequency scale." In other words, in the presence of motion, humans are more sensitive to the wavelet decomposition coefficients at the coarser scales than at the finer scales. We enforce this perceptual phenomenon in our STRRED indices by computing the SRRED and TRRED indices in the coarsest passband of the steerable pyramid decomposition of the frames and the frame differences, respectively. We also observed empirically that indices computed at these scales gave the best performance.

VI. Results and Discussion

We now present the results of the correlation analysis of the STRRED indices on the LIVE video quality assessment database [5]. The LIVE video quality assessment database contains ten reference videos and 150 distorted videos spanning four categories of distortions, 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. Six videos contain 250 frames at 25 f/s, one video contains 217 frames at 25 f/s, and three videos contain 500 frames at 50 f/s.

A. Implementation Details

The luminance frames in the video sequence, as well as the luminance frame differences, are subjected to a multiscale multiorientation wavelet decomposition using steerable pyramids [19]. The decomposition is performed at three scales and six orientations. Every subband is partitioned into nonoverlapping blocks, each of size 3×3. The value of the neural noise

TABLE II
RELATION BETWEEN T, S, AND OF

Sequence	1	2	3	4	5	6	7	8	9	10
$\mathbb{E}[S]$	1.00	0.99	0.97	0.97	0.99	1.01	0.99	1.00	0.99	1.00
$\mathbb{E}[S\|T>1]$	1.79	2.04	2.01	1.38	1.95	2.78	2.17	1.60	2.44	1.42
$\mathbb{E}[OF]$	2.01	2.98	0.78	4.03	0.83	2.58	2.21	1.33	0.97	1.04
$\mathbb{E}[OF\|T>1]$	3.61	4.95	1.64	4.05	1.15	2.77	3.14	1.41	1.10	1.06

TABLE V
SROCC BETWEEN SINGLE-NUMBER STRRED INDICES AND LIVE VIDEO QUALITY ASSESSMENT DATABASE SCORES

Distortion Type	STRRED_4^1	$\text{SRRED}_4^{1/2}$	$\text{TRRED}_4^{1/2}$
Wireless errors	0.7208	0.6066	0.5863
IP errors	0.5075	0.3851	0.5279
H.264	0.7197	0.4441	0.6737
MPEG	0.7247	0.7540	0.4363
Overall	0.7319	0.5961	0.5870
No. of scalars per frame	1	1/2	1/2

TABLE VII
LCC BETWEEN SINGLE-NUMBER STRRED INDICES AND LIVE VIDEO QUALITY ASSESSMENT DATABASE SCORES

Distortion Type	STRRED_4^1	$\text{SRRED}_4^{1/2}$	$\text{TRRED}_4^{1/2}$
Wireless errors	0.7051	0.6218	0.5991
IP errors	0.5453	0.4148	0.5096
H.264	0.7242	0.4270	0.7058
MPEG	0.7490	0.7605	0.4292
Overall	0.7264	0.6057	0.5741
No.of scalars per frame	1	1/2	1/2

variance for both spatial and temporal data is chosen to be 0.1, i.e., $\sigma_W^2 = \sigma_Z^2 = 0.1$. Note that similar values of the neural noise variance were chosen in [4] and [21].

B. Subjective Evaluation

The SRRED, TRRED, and STRRED indices are evaluated against subjective quality scores on the LIVE video quality assessment database. As mentioned earlier, the wavelet coefficients in the coarsest passband (for both the decomposition of the frames and the decomposition of the frame differences) yield the best performance, and all the performance results reported in this section are based on these wavelet coefficients. Further, we compare the performance of the STRRED indices evaluated in different orientations at the coarsest scale in Table III. We use the Spearman rank order correlation coefficient (SROCC) between the subjective scores and the quality indices to compare the relative performances between the orientations. $\text{STRRED}_4^{M_4}$ yields the best performance among different orientations and this corresponds to the vertically oriented subband. In the rest of this section, we present detailed comparisons of the SRRED, TRRED, and STRRED indices evaluated in the vertically oriented subband at the coarsest scale against other VQA algorithms.

Table IV contains a detailed comparison of the SROCC of those RRED indices that operate at a high data rate against PSNR, multiscale (MS)-SSIM (computed for every frame and averaged across frames) [8], VQM [15], and the MOVIE index [1] against human opinion scores available with the LIVE VQA database [5].

The results in Table IV show that the STRRED algorithms using $L/576$ scalars perform as good as some of the best FR VQA algorithms, such as the MOVIE index. A similar behavior is also observed for the SROCCs of the single-number algorithms presented in Table V. We recall that these single-number algorithms require just a single number per frame from the reference or distorted video for quality computation. It is clear from the results that even the single-number algorithms significantly outperform PSNR, which is, in fact, an FR quality index. The performance of the single-number algorithms is almost at par with some of the FR VQA algorithms, such as MS-SSIM. It is interesting that while the individual performances of the single-number versions of the SRRED and TRRED indices are not outstanding, it is their combination that renders the algorithm effective. It appears that SRRED and TRRED seem to be capturing complementary distortions and that their combination makes them highly competitive.

We observe that for MPEG distortions, the single-number RRED indices appear to perform better than the ones that use more information from the reference. This occurs because the pooling strategies of the single-number algorithms and the ones that use more reference information are different. In principle, the pooling strategy employed by the single-number algorithms can also be used by the algorithms that operate using more information from the reference. However, their overall performance would be poorer if such a pooling strategy were to be performed. We would like to clarify that while the algorithms that use more information from the reference can use the pooling strategy of the single-number algorithms, the reverse is not possible. The single-number algorithms do not have enough information to employ any other strategy since they are supplied with just one scalar.

We report the results of linear correlation coefficient (LCC) scores in Tables VI and VII. We use nonlinearity on the objective scores before computing the LCC. For the STRRED indices, we use the nonlinearity given by

$$\text{Quality}(x) = \beta_1 \log(1 + \beta_2 x).$$

The fit between the subjective and objective scores used to compute the LCC for the STRRED indices is shown in Fig. 4. The LCC scores also follow similar trends as compared with the SROCC scores.

While in Tables IV and V we discuss the performance of STRRED indices that operate at two extremes of the amount of information from the reference, in Table VIII, we show the

TABLE III

SROCC BETWEEN STRRED INDICES AT DIFFERENT ORIENTATIONS IN THE COARSEST SCALE AND LIVE VIDEO QUALITY ASSESSMENT DATABASE SCORES

Index	$\text{STRRED}_2^{M_2}$	$\text{STRRED}_3^{M_3}$	$\text{STRRED}_4^{M_4}$	$\text{STRRED}_5^{M_5}$	$\text{STRRED}_6^{M_6}$	$\text{STRRED}_7^{M_7}$
SROCC	0.7508	0.7958	0.8007	0.7674	0.7330	0.7187

TABLE IV

SROCC BETWEEN STRRED INDICES, PSNR, MS-SSIM, VQM, MOVIE, AND LIVE VIDEO QUALITY ASSESSMENT DATABASE SCORES

Distortion Type	$\text{STRRED}_4^{M_4}$	$\text{SRRED}_4^{M_4}$	$\text{TRRED}_4^{M_4}$	PSNR	MS-SSIM	VQM	MOVIE
Wireless errors	0.7857	0.7925	0.7765	0.6574	0.7289	0.7214	0.8114
IP errors	0.7722	0.7624	0.7513	0.4167	0.6534	0.6383	0.7192
H.264	0.8193	0.7542	0.8189	0.4585	0.7313	0.6520	0.7797
MPEG	0.7193	0.7249	0.5879	0.3862	0.6684	0.7810	0.8170
Overall	0.8007	0.7592	0.7802	0.5398	0.7364	0.7026	0.8055
No. of scalars per frame	L/576	L/1152	L/1152	L	L	L/25	L

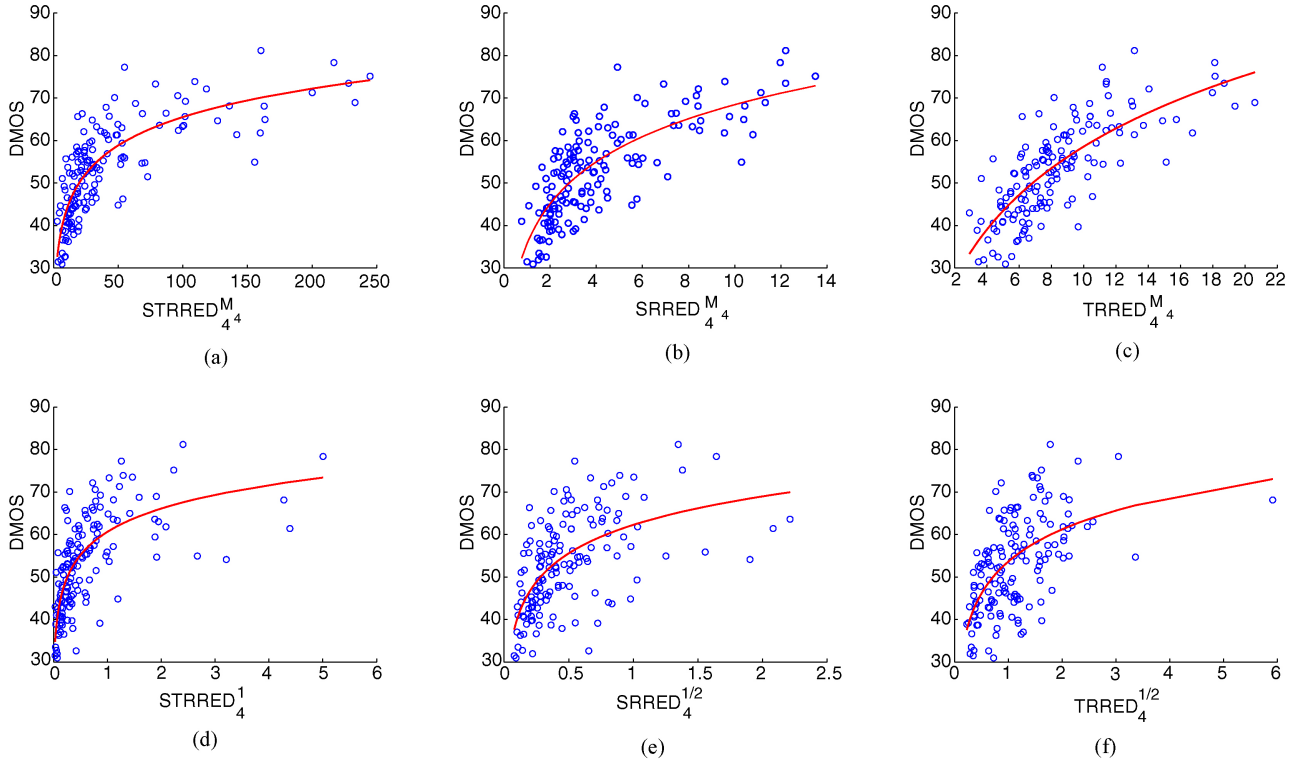

Fig. 6. Nonlinear fit used to compute linear correlation coefficient on the overall LIVE video quality assessment database. (a) $\text{STRRED}_4^{M_4}$. (b) $\text{SRRED}_4^{M_4}$. (c) $\text{TRRED}_4^{M_4}$. (d) STRRED_4^{1}. (e) $\text{SRRED}_4^{1/2}$. (f) $\text{TRRED}_4^{1/2}$.

TABLE VI

LCC BETWEEN STRRED INDICES, PSNR, MS-SSIM, VQM, MOVIE, AND LIVE VIDEO QUALITY ASSESSMENT DATABASE SCORES

Distortion Type	$\text{STRRED}_4^{M_4}$	$\text{SRRED}_4^{M_4}$	$\text{TRRED}_4^{M_4}$	PSNR	MS-SSIM	VQM	MOVIE
Wireless errors	0.8039	0.8067	0.7726	0.6695	0.7157	0.7325	0.8371
IP errors	0.8020	0.8033	0.7619	0.4689	0.7267	0.6480	0.7383
H.264	0.8228	0.7462	0.8324	0.5330	0.7020	0.6459	0.7920
MPEG	0.7467	0.7281	0.5998	0.3986	0.6640	0.7860	0.8252
Overall	0.8062	0.7764	0.7743	0.5604	0.7379	0.7326	0.8217
No. of scalars per frame	L/576	L/1152	L/1152	L	L	L/25	L

TABLE VIII
VARIATION OF OVERALL SROCC WITH THE AMOUNT OF INFORMATION REQUIRED FROM THE REFERENCE FOR THE STRRED INDICES

No. of scalars per frame	SROCC
L'	0.8007
0.250 L'	0.8056
0.167 L'	0.8056
0.083 L'	0.8015
0.042 L'	0.7992
0.021 L'	0.7930
0.014 L'	0.7838
0.010 L'	0.7660
0.002 L'	0.7319

variation in the performance as we increase the amount of information. The SROCC computed on the overall database is shown for the STRRED indices. The amount of information is reduced by computing and sending partial sums of the scaled local entropies instead of the scaled entropies at all the locations as in [4]. The table shows that the algorithm that requires $L/27648$ scalars from the reference achieves a performance that is almost as good as the algorithm that requires $L/576$. We denote $L' = L/576$, where L is the number of pixels in a frame of the video. Note that $0.002L'$ corresponds to the single-number algorithm that requires just a single scalar from the reference per frame. The minor increase in the performance of the STRRED indices with decrease of information between rows 1 and 2 in Table VIII is due to the difference between computing the sum of the absolute differences of partial sums and computing the sum of the absolute differences of the scaled entropies at different locations. This phenomenon was also explained earlier in this section with regard to the superior performance of the single-number algorithms for MPEG distortions over the algorithms that operate at higher data rates of reference information.

C. Computational Complexity and Runtimes

The computational complexity of the STRRED indices closely follows that of the RRED indices [4]. Both SRRED and TRRED indices require $O(N(\log N + M^2))$ operations per frame per subband, where $\sqrt{M} \times \sqrt{M}$ is size of a block during processing of each subband. We refer the reader to [4] for a detailed calculation. Technically, the TRRED indices require a differencing step prior to computation of the wavelet transform, which requires N more operations and is consumed in the order notation. Thus, the overall complexity of the STRRED indices may be written as $O(FN(\log N+M^2))$, where F is the number of frames.

We also report runtime results of experiments conducted on an Intel Pentium processor with 2 GB RAM and 3.4 GHz speed using MATLAB (R2008a) without code optimization. For a sequence of length 250 frames, the multiscale multiorientation transform using steerable pyramids takes around 215 s, while the computation of the scaled entropy information that is finally differenced takes 4–6 s for any (reference or distorted) video. Note that the steerable pyramid decomposition time reported above is the time taken to compute the decomposition into 26 subbands (at four scales and six orientations). By operating the STRRED algorithms in only one of the subbands, it is possible to reduce the time required for this step. Despite this, it appears that the steerable pyramid decomposition is the bottleneck and improving the efficiency of this step will help improve the overall runtime of the algorithm.

VII. CONCLUSION

We developed a family of RR VQA algorithms that vary in the amount of reference information required for quality computation. These algorithms were based on statistical models for videos in both spatial and temporal domains and computed the differences in the amount of information between the reference and distorted videos to measure quality. While the algorithms with more information from the reference video approached the performance of full reference VQA algorithms, the single-number algorithm outperformed PSNR. Depending on the application and amount of reference information that can be afforded, a user could pick an algorithm from this class for automatic prediction of video quality. The computation of the wavelet transform was a bottleneck in terms of the time complexity of the algorithm. The idea developed in this paper could be useful in other video representations with well-behaved statistical models and faster computational times that will enable faster implementations of the algorithm.

We studied the use of frame differences between adjacent frames to design the STRRED indices here. One interesting future direction is to explore the use of differences between frames that are spaced farther apart and to develop good statistical models for these in order to be able to apply them to VQA algorithms. The question of whether such differences are relevant for video QA also needs to be addressed.

REFERENCES

[1] 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.
[2] Z. Wang and Q. Li, "Video quality assessment using a statistical model of human visual speed perception," *J. Opt. Soc. Am. A Opt. Image Sci. Vision*, vol. 24, no. 12, pp. B61–B69, Dec. 2007.
[3] S. Roth and M. Black, "On the spatial statistics of optical flow," *Int. J. Comput. Vision*, vol. 74, no. 1, pp. 33–50, Aug. 2007.
[4] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 517–526, Feb. 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] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. Broadcast.*, vol. 57, no. 2, pp. 165–182, Feb. 2011.
[7] S. Winkler, *Digital Video Quality: Vision Models and Metrics*. New York: Wiley, 2005.
[8] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals Syst. Comput.*, Nov. 2003, pp. 1398–1402.
[9] Sarnoff Corporation. (2003). *JNDmetrix Technology* [Online]. Available: http://www.sarnoff.com/productsservices/videovision/jndmetrix/downloads.asp
[10] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[11] A. B. Watson, J. Hu, and J. F. McGowan, III, "Digital video quality metric based on human vision," *J. Electron. Imag.*, vol. 10, no. 1, pp. 20–29, Jan. 2001.
[12] A. P. Hekstra, J. G. Beerends, D. Ledermann, F. E. de Caluwe, S. Kohler, R. H. Koenen, S. Rihs, M. Ehrsam, and D. Schlauss, "PVQM: A perceptual video quality measure," *Signal Process.: Image Commun.*, vol. 17, no. 10, pp. 781–798, Nov. 2002.
[13] M. Barkowsky, J. Bialkowski, B. Eskoer, R. Bitto, and A. Kaup, "Temporal trajectory aware video quality measure," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 266–279, Apr. 2009.
[14] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Considering temporal variations of spatial visual distortions in video quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 253–265, Apr. 2009.
[15] 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.
[16] M. Masry, S. S. Hemami, and Y. Sermadevi, "A scalable wavelet-based video distortion metric and applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 2, pp. 260–273, Feb. 2006.
[17] I. P. Gunawan and M. Ghanbari, "Reduced-reference video quality assessment using discriminative harmonic strength with motion consideration," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 71–83, Jan. 2008.
[18] M. Rohani, A. N. Avanaki, S. Nader-Esfahani, and M. Bashirpour, "A reduced reference video quality assessment method based on the human motion perception," in *Proc. 5th Int. Symp. Telecommun.*, 2010, pp. 831–835.
[19] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 1995, pp. 444–447.
[20] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images," in *Proc. Adv. Neural Inform. Process. Syst.*, vol. 12. May 2000, pp. 64–68.
[21] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.
[22] P. Thompson, "Perceived rate of movement depends on contrast," *Vision Res.*, vol. 22, no. 3, pp. 377–380, 1982.
[23] P. Thompson, K. Brooks, and S. T. Hammett, "Speed can go up as well as down at low contrast: Implications for models of motion perception," *Vision Res.*, vol. 46, nos. 6–7, pp. 782–786, 2006.
[24] E. P. Simoncelli and D. J. Heeger, "A model of neuronal responses in visual area MT," *Vision Res.*, vol. 38, no. 5, pp. 743–761, Mar. 1998.
[25] D. B. Hamilton, D. G. Albrecht, and W. S. Geisler, "Visual cortical receptive fields in monkey and cat: Spatial and temporal phase transfer function," *Vision Res.*, vol. 29, no. 10, pp. 1285–1308, 1989.
[26] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, Jun. 2010, pp. 2432–2439.
[27] D. C. Burr and J. Ross, "Contrast sensitivity at high velocities," *Vision Res.*, vol. 23, no. 4, pp. 3567–3569, 1982.

Rajiv Soundararajan received the B.E. (Hons.) degree in electrical and electronics engineering from the Birla Institute of Technology and Science, Pilani, Rajasthan, India, in 2006, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, in 2008 and 2012, respectively.

He is currently with Qualcomm Research India, Bangalore, India. His current research interests include statistical image and video processing and information theory, and more specifically, compression, quality assessment, and computer vision applications.

Alan C. Bovik (F'96) is the Curry/Cullen Trust Endowed Chair Professor with the University of Texas at Austin, Austin, where he is currently the Director of the Laboratory for Image and Video Engineering. He is a Faculty Member with the Department of Electrical and Computer Engineering and the Center for Perceptual Systems, Institute for Neuroscience. He holds two U.S. patents. His current research interests include image and video processing, computational vision, and visual perception. He has published almost 600 technical articles in these areas. His several books include the recent companion volumes *The Essential Guides to Image and Video Processing* (New York: Academic, 2009).

Dr. Bovik was named the SPIE/IS&T Imaging Scientist of the Year in 2011. He was a recipient of a number of major awards from the IEEE Signal Processing Society, including the Best Paper Award in 2009, the Education Award in 2007, the Technical Achievement Award in 2005, and the Meritorious Service Award in 1998. He received the Hocott Award for Distinguished Engineering Research from the University of Texas at Austin, the Distinguished Alumni Award from the University of Illinois at Urbana-Champaign, Urbana, in 2008, the IEEE Third Millennium Medal in 2000, and two Journal Paper Awards from the International Pattern Recognition Society in 1988 and 1993. He is a fellow of the Optical Society of America, the Society of Photo-Optical and Instrumentation Engineers, and the American Institute of Medical and Biomedical Engineering. He has been involved in many professional society activities, including serving as a member of the Board of Governors and the IEEE Signal Processing Society from 1996 to 1998, a Co-Founder and the Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002, an Editorial Board Member of the PROCEEDINGS OF THE IEEE from 1998 to 2004, the Series Editor of *Image, Video, and Multimedia Processing* since 2003, and the Founding General Chairman of the First IEEE International Conference on Image Processing, Austin, in November 1994. He is a Registered Professional Engineer in Texas and a frequent Consultant to legal, industrial, and academic institutions.