Video Quality Pooling Adaptive to Perceptual Distortion Severity

Jincheol Park, Kalpana Seshadrinathan, *Member, IEEE*, Sanghoon Lee, *Senior Member, IEEE*, and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*— It is generally recognized that severe video distortions that are transient in space and/or time have a large effect on overall perceived video quality. In order to understand this phenomena, we study the distribution of spatio-temporally local quality scores obtained from several video quality assessment (VQA) algorithms on videos suffering from compression and lossy transmission over communication channels. We propose a content adaptive spatial and temporal pooling strategy based on the observed distribution. Our method adaptively emphasizes "worst" scores along both the spatial and temporal dimensions of a video sequence and also considers the perceptual effect of large-area cohesive motion flow such as egomotion. We demonstrate the efficacy of the method by testing it using three different VQA algorithms on the LIVE Video Quality database and the EPFL-PoliMI video quality database.**

***Index Terms*— Egomotion, perceptually influential distortions, pooling, video quality assessment, VQPooling.**

I. Introduction

VIDEO quality assessment (VQA) deals with predicting the perceptual quality of a video, *i.e.*, the quality of a video as judged by an average human observer. A large body of work on VQA exists in the literature and much of the work has focused on full reference VQA [1]– [3]. Full reference VQA algorithms estimate the quality of a test video, assuming the availability of a pristine reference video that was used to create the test video. Most full reference VQA algorithms estimate the quality of a video over local spatio-temporal regions [4]–[6]. The sizes of the regions of support used by different algorithms vary along both the spatial and temporal dimensions. Some VQA algorithms operate on a frame-by-frame basis, while others consider several frames of the video sequence when predicting local, spatio-temporal quality [1]– [13]. These spatio-temporal local quality indices are then combined into a single global quality index of the entire video, which predicts an average human observers' quality opinion of the video.

There is no perfect understanding of the way that human observers combine local spatio-temporal local impressions of video quality to obtain judgments of overall video quality although several researchers have studied this question [1]. The simple strategy of taking the average, or sample mean, of the local quality indices over both spatial and temporal coordinates to obtain a single global quality index has been employed in the construction of many VQA algorithms [2]– [7]. Several other models have also been proposed regarding how spatial and temporal local quality indices should be combined into a single global quality index [8]–[10].

Recently, several studies have focused on developing temporal pooling mechanisms that combine per-frame quality indices into an overall quality index. The approach of [11] considers both short-term and long-term temporal variations of spatial distortions using a wavelet-based quality assessment (WQA) model to develop temporal pooling mechanisms [12]. Forgiveness and negative peak duration neglect effects, where overall ratings were greatly influenced by the single most severe event while the duration of the event was neglected, were reported using data gathered using a single stimulus continuous quality evaluation (SSCQE) paradigm in [14], [15]. The degree of smoothness of subjective time-varying quality scores was observed and modeled in [16]. A hysteresis effect on the subjective judgment of video quality was observed in a recent study of time-varying video quality [17]. Temporal pooling of quality scores of networked video in packet loss situations was studied by the authors of [18], who determined that perceived temporal quality degradations are predominately determined by the duration over which each frame is displayed.

In [19]– [22], a foveal peak signal-to-noise ratio (FPSNR) used known or predicted fixation locations, along with a spatial weighting mechanism to predict visual quality. However, predicting human fixations remains a difficult and unsolved problem. More recently, the authors of [23] defined an attention map encapsulating various factors such as color, orientation, motion, etc. when combining local quality scores. After sorting the attention map values in descending order, twenty percent of

Manuscript received April 26, 2012; revised August 30, 2012; accepted September 2, 2012. Date of publication September 18, 2012; date of current version January 10, 2013. This work was supported in part by the Yonsei University Institute of TMS Information Technology, Korea, through the Brain Korea 21 Program, the Technology Innovation Program Standardization of 3-D Human Factors and 3-D Medical Application Services funded by the Ministry of Knowledge Economy (MKE) of Korea under Grant 10042402, and the Ministry of Knowledge Economy, Korea, through the Information Technology Research Center Support Program supervised by the National IT Industry Promotion Agency under Grant NIPA-2012-H0301-12-1008. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alex C. Kot.

J. Park and S. Lee are with the Wireless Network Laboratory, Department of Electrical and Electronics Engineering, Yonsei University, Seoul 120-749, Korea (e-mail: dewofdawn@yonsei.ac.kr; slee@yonsei.ac.kr).

K. Seshadrinathan is with Intel Corporation, Santa Clara, CA 94086 USA (e-mail: kalpana.seshadrinathan@intel.com).

A. C. Bovik is with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, University of Texas, Austin, TX 78712-1084 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/TIP.2012.2219551

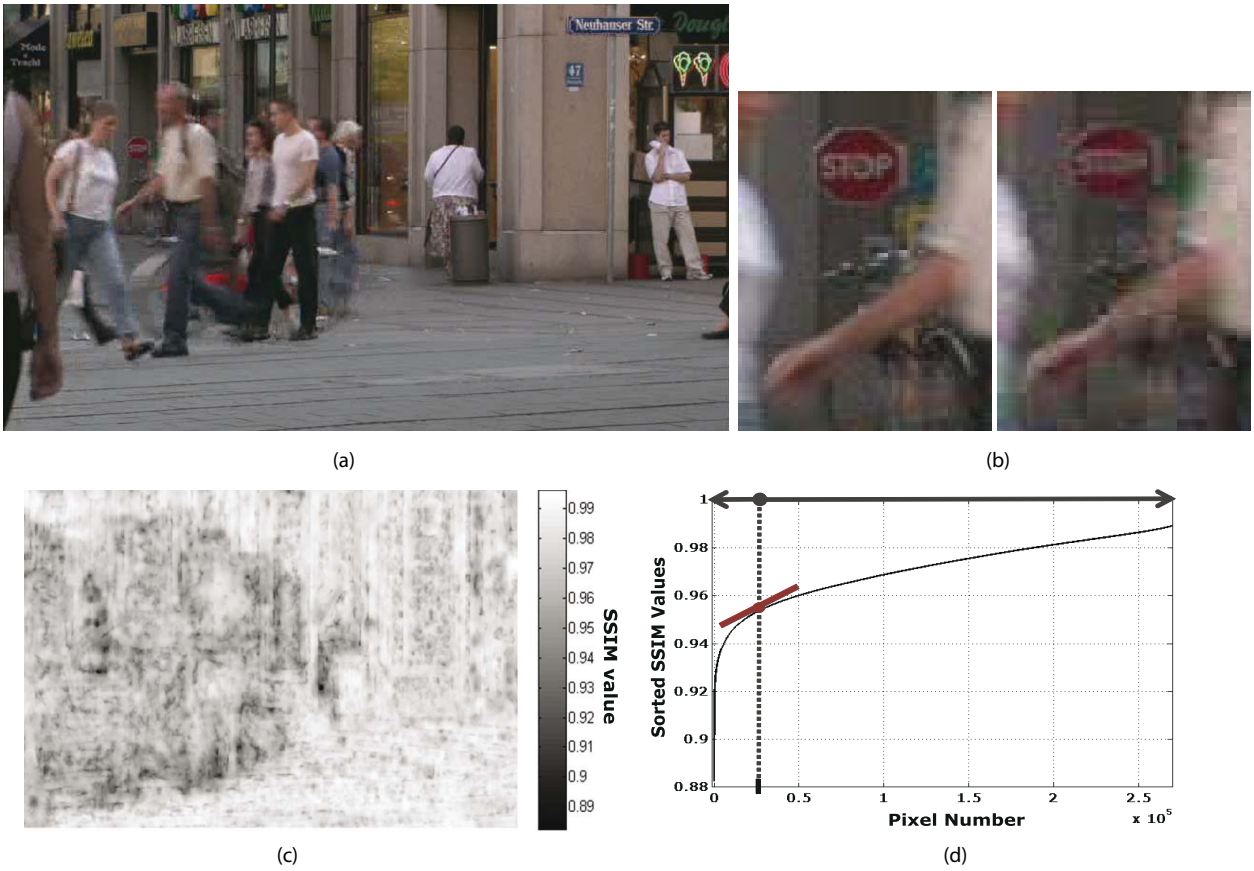

Fig. 1. Typical distribution of spatio-temporal video quality. (a) 81st frame of the *Pedestrian Area 16* (pa16) sequence in the LIVE Video Quality database distorted by MPEG-2 compression. (b) Zoomed-in view of (a), showing a region suffering from severe encoding distortions (right) side-by side with the reference region (left). (c) Quality index map obtained by SSIM expressing local spatial quality at each pixel. (d) Curve showing the SSIM values in (c), sorted in ascending order.

the blocks having the highest attention map values are selected as candidates for spatio-temporal pooling.

It has been hypothesized that the worst local quality scores (both spatial and temporal) affect the overall subjective perception of quality more significantly than do the regions of good quality [5]–[28]. This observation was used to devise a pooling mechanism that sorts and weights a fixed percentile of poor quality scores higher than the remaining scores [5]. Percentile pooling using the lowest $p\%$ of quality scores to predict the final score was also studied in [24] as a means of emphasizing the stronger influence of more annoying, lower quality regions of still images. The mean time between failures (MTBF), representing how often noticeable visual errors are observed, was used as an indicator of subjective quality perception in [27]. Machine learning methods were explored to determine the impact of spatial and temporal factors as well as their interaction on overall video quality in [28].

We propose a more comprehensive spatial and temporal pooling model that is based on the distributions of local spatio-temporal quality scores from objective VQA algorithms and that relies on both perceptual and behavioral models of distortion perception. Our model adapts to the video content, and parameters used in the spatio-temporal pooling are determined for each video sequence based on the distribution of the spatio-temporal quality scores and the presence of large coherent motion (such as egomotion) in the video. The distribution of quality scores in a video frame can vary considerably, and can significantly impact human judgments of overall video quality. We hypothesize that it is essential to adaptively extract the scores and weights that are used for pooling by systematically considering the distribution of the local spatio-temporal quality scores. We also observe that the distribution of local spatio-temporal quality scores are affected by the presence or absence of large, cohesive motion regions in a video frame and incorporate this effect into our pooling strategy. Such motion can arise from egomotion, i.e., optical flow or motion fields induced by camera motion.

We evaluate the proposed approach, which we dub Video Quality Pooling (VQPooling), on quality maps delivered by different VQA algorithms on the distorted videos in the Laboratory for Image and Video Engineering (LIVE) Video Quality Database [29] and the EPFL-PoliMI video quality database [30]. The results show clear improvement in the performance of these algorithms as compared to traditional mean-based or percentile-based pooling methods. The work reported here builds on our preliminary work in [32]. We perform a rigorous performance evaluation of VQPooling on additional VQ databases and also perform a statistical performance analysis against traditional pooling methods [32].

(a) (b) (c)

Fig. 2. Comparison of the distribution of local spatio-temporal quality scores for different distortion types. (a) Compression distortion only. (b) Compression and distortions due to errors in communication channel. (c) Curve showing the local spatio-temporal SSIM scores of (a) and (b) sorted in ascending order.

II. Spatio-Temporal Characteristics of Video Distortion

A. Distribution of Spatio-Temporal Video Quality

Video compression and lossy transmission processes do not cause uniform distortions over a video frame and instead, cause non-uniform distortions that vary in the level of perceptual annoyance they cause over space and time. In particular, severe and highly annoying distortions that occur locally in space or time heavily influence an observers judgment of quality [14], [5]. As depicted in Fig. 1(a), video frames can be viewed as being composed of large areas of smooth, low-frequency spatial variations, broken by sharp edges and textures occurring between. Natural video frames exhibit this property, which forms the basis of various natural scene statistics models. Indeed, the reduced sensitivity of humans to local high frequencies is used in video compression. High spatial frequency coefficients are more severely quantized than low frequency coefficients in typical video coding schemes such as the MPEG standards. Hence, encoding distortions such as blur and loss of detail are often more severe in such regions of high spatial activity, rather than in flat regions [31], as shown in Fig. 1(b) and (c). Further, certain distortions such as blocking tend to occur uniformly throughout a frame of a video. However, the perceptual visibility of blocking distortions varies based on scene content. Blocking is more visible in smooth regions, while heavy spatial activity can mask blocking artifacts, rendering them less visible or annoying [33].

The phenomenon described above applies to the spatial distribution of quality scores in a single intra-coded frame of the distorted video. A similar reasoning is applied to predictively coded video frames due to the characteristics of natural videos and natural video distortions along the temporal domain. Along the temporal dimension there is a high correlation between neighboring frames. Typical video compression algorithms utilize motion compensated block differences across frames to reduce temporal correlations, thereby enhancing compressibility. The prediction errors over smooth and static regions of the video is often small. However, large prediction errors are produced around the borders of moving objects, often resulting in local regions exhibiting severe distortion [31], [34]. Thus, predicted frames also suffer from small areas that are severely distorted, interspersed among larger areas of better quality.

We performed experiments using SSIM as an indicator of spatio-temporally local quality and observed the distributions of these scores using distorted videos in the LIVE Video Quality Database. We found that if the SSIM scores for a frame of the video are sorted in rank order, the sorted quality scores tend to follow a curve as depicted in Fig. 1(d) that saturates at higher quality scores. This curve shape contains a saturating portion that arises from regions suffering from very low degrees of quality degradation. The steeply increasing portion of the curve corresponds to regions of severe quality degradation.

The distribution of objective quality scores in video frames that suffer from distortions introduced by lossy transmission of video depends on the nature of the lossy channel. Videos are compressed by video coding algorithms before being transmitted over networks. Channel distortions typically produce much more significant distortions than do compression distortions [35], [36]. We found that videos suffering from distortion due to network errors also exhibit a similar distribution as compression distortions. Further, we found that severe quality degradations were introduced by the channel errors resulting in a steeper slope of the sorted quality scores. Fig. 2(a) and (b) illustrate this effect for identically compressed video sequences. However, the video in Fig. 2(b) additionally suffers from channel distortion, resulting in a steeper slope due to more severe quality degradation.

We hypothesize that a video sequence that has undergone distortion through destructive processes such as compression and lossy network transmission will typically produce a distribution of objective quality scores in a frame similar to the shape in Fig. 1. In our experiments, we have observed this to be true on every video in the LIVE Video Quality Database, the VQEG FRTV Phase I database and the EPFL PoliMI database. We have also observed similar distributions of objective quality scores using a variety of objective VQA algorithms such as MSE, SSIM and MOVIE.

However, we have not yet been able to find any consistent temporal models that can account for the frame level objective quality scores obtained by spatial pooling. We believe that this follows since temporal video quality can fluctuate considerably with video content, error propagation due to motion compensated video encoding and channel induced distortions [37], [38]. However, temporal regions of severe quality degradation do impact visual quality [11], [14], [26] and we account

(a) (b)

Fig. 3. Illustration of weakness of the sample mean as a spatial pooling method on two distorted versions of the *Mobile & Calendar* sequence from the LIVE Video Quality database. (a) 188th frame of mc2 (mean SSIM = 0.8495, mean MSE = 119.92, and mean spatial MOVIE = 0.9603). (b) 188th frame of mc12 (mean SSIM = 0.8339, mean MSE = 123.41, and mean spatial MOVIE = 0.9584).

for this effect by clustering temporal video quality scores into regions of high and low quality.

We utilize these observations to develop an approach for pooling local spatial and temporal quality indices into a single quality index in Section II-B.

B. Proposed Approach

We have discussed the hypothesis that the most severe impairments substantially affect the overall subjective perception of quality. This is nicely illustrated in Fig. 3, where the video frame in Fig. 3(a) suffers from severe localized degradations, while the distribution of distortions in Fig. 3(b) is more uniform. However, even though the overall quality of Fig. 3(a) is worse than Fig. 3(b), the quality scores delivered by SSIM, MSE and MOVIE on the distorted video depicted in Fig. 3(b) are worse than those on the distorted video in Fig. 3(a) using spatial mean pooling. The problem of extracting such influential "worst quality" scores in a content adaptive manner has not been adequately studied. We explain a natural approach to accomplish this.

As discussed in Section II-A, video quality scores sorted in ascending order typically exhibit a curve with a saturating tendency, with lower quality scores occurring in the steeply increasing region of the curve and higher quality scores occurring in the saturated region. Thus, we may view the problem of extracting perceptually influential poor quality regions of the 2-dimensional quality map corresponding to severe distortions as a classification problem on the saturating curve. In other words, the goal is to separate the increasing and saturating regions of the 1-dimensional curve in Fig. 1. Following this line of reasoning, we construct a spatial pooling model that we dub *spatial VQPooling* [32] (VQ = Video Quality). It determines the increasing region of the saturating curve that maps the worse quality scores, then emphasizes them when determining the overall video quality score. Later we also introduce temporal and motion-based aspects of the VQPooling approach.

We also observe that the distribution of local spatio-temporal quality scores are affected by the presence or absence of large, cohesive motion fields such as egomotion in a video frame and incorporate this effect into our pooling strategy. Ego-motion refers to the presence of optical flow or motion fields induced by camera motion. Egomotion induces cohesive, often large velocity fields in videos that may mask local distortions [9], [39], [40], making them less visible. This is powerfully demonstrated by the "silencing" effect demonstrated in [40], where local image changes of highly diverse types (size, shape, flicker, hue) are "silenced" or masked by larger collective movements. Likewise, quality scores may be affected by such velocity fields by modifying their sensitivity to local spatial or temporal distortions. Any sufficiently large cohesive motion field may produce similar effects. We use different criteria to extract perceptually significant low quality regions of the video for frames containing large, coherent motion fields to account for these effects. Details are described in Section III-B.

Along the temporal domain, frame level quality scores are divided into two groups of higher and lower quality using a clustering algorithm. Perceptually influential worse quality scores are emphasized in the overall quality index. Moreover, we found that the magnitude of the difference between the qualities in the higher and lower quality groups affects the impression of overall quality. The quality scores of the higher group are adaptively weighted using the difference of the scores between the higher and lower quality groups in a temporal pooling model that we name *temporal VQPooling* [32]. Details are described in Section III-C.

III. Content Adaptive Spatial and Temporal Pooling

A. Egomotion Detection

To detect egomotion or similar large, cohesive motion fields in a frame, we deploy a simple method that uses the first and second-order statistics (means and standard deviations) of the magnitudes of the computed motion vectors (MVs). When there is no camera motion, local portions of frames may contain diverse and disparate MVs, while "background" regions have zero or small MVs. This results in the mean magnitude of the MVs in the frame being very low, while the standard deviation is relatively high. However, if there

is camera motion or other large, collective motion, then large percentages of MVs will be similar with high mean displacement values, while the standard deviation of the MVs becomes relatively small. MVs are computed in VQPooling using a standard full search motion estimation algorithm on 16×16 macroblocks (MBs) [41]. The decision regarding the presence of egomotion is made based on a simple predicate. When the standard deviation of the MV's is larger than the mean, then no significant camera movement or other large, collective motion is deemed present; otherwise egomotion or similar large, cohesive motion field is deemed to be present.

B. Content Adaptive Spatial Pooling

One contribution of our work is that we describe a method of extracting regions of highly influential poor quality in an adaptive manner for each video frame. Our classification of saturated and increasing regions of the curve takes into account the distribution of the quality scores. We perform this classification based on the slope of the quality score curve.

Let z denote an index into the set of sorted scores and let $Q_f(z)$ denote the sorted spatially local quality scores obtained using a VQA algorithm on a frame f. Note that the VQA algorithm may operate using a single frame or a neighborhood of frames to deliver the quality score $Q_f(z)$. Thus, $Q_f(z)$ is the z^{th} lowest quality score. The discrete derivative of $Q_f(z)$, denoted $Q'_f(z)$, estimates the slope of this curve:

$$Q'_f(z) \approx \frac{\bar{Q}_f(z+\Delta) - \bar{Q}_f(z)}{\Delta} \cdot N_s \tag{1}$$

where N_s is the number of quality scores in a frame and $\bar{Q} = \frac{Q_f(z) - Q_f^{Min}}{Q_f^{Max} - Q_f^{Min}}$, where Q_f^{Min} and Q_f^{Max} are the minimum and maximum scores in the f^{th} frame, respectively. Thus $0 \leq \bar{Q}_f \leq 1$, regardless of the quality index used.

A slope criterion is used to classify quality scores into increasing and saturated regions. Let t_f be a threshold that is determined based on the degree of estimated egomotion or other collective motion in the frame. Let μ_f^v and σ_f^v denote the mean and standard deviation of the magnitudes of MVs of the f^{th} frame respectively. The presence of egomotion, or similar global motion, is decided using the coefficient of variation (CoV) of the magnitude of the MVs. The CoV is defined as the ratio of the standard deviation to the mean. A frame is considered to be moving when the CoV is lower than or equal to 1, which corresponds to the standard deviation of the motion vectors being smaller than their mean value. This simple threshold of 1 provides very good results for identifying frames with egomotion. The slope threshold t_f is determined based on the presence or absence of ego-motion according to:

$$t_f = \begin{cases} t_M, & \frac{\sigma_f^v}{\mu_f^v} < 1 \\ t_S, & \frac{\sigma_f^v}{\mu_f^v} \geq 1. \end{cases} \tag{2}$$

The constants t_M and t_S therefore become the slope criterion for frames with and without egomotion, respectively. Since frames without egomotion typically contain smaller regions suffering from severe quality degradation, we fix $t_S > t_M$. Further details on the choice of these parameters are provided in Section IV. The larger slope criterion that is applied to frames without large coherent motion better separates the influential low quality scores in these frames.

The *increasing region* of the sorted quality score curve is the set

$$P_t = \{z : Q'_f(z) < t_f\} \tag{3}$$

with complement

$$P_t^C = \{z : Q'f(z) \geq t_f\} \tag{4}$$

which is the quality saturation region of the sorted quality score curve.

A frame level quality index s_f for frame f is then computed:

$$s_f = \frac{\sum_{z \in P_t} Q_f(z) + r \cdot \sum_{z \in P^c{}_t} Q_f(z)}{|P_t| + r \cdot |P^c{}_t|} \tag{5}$$

where $|P_t|$ denotes the cardinality of P_t, and $r \ll 1$ is a small multiplier that is used to account for the reduced perceptual contribution of the scores in P_t^c to the overall quality of the video.

C. Content Adaptive Temporal Pooling

The spatial pooling strategy described in Section III-B produces frame level quality indices s_f that are determined in a content adaptive manner. We now perform content adaptive temporal pooling to aggregate these frame level quality indices into an overall quality index for the entire video.

To perform temporal pooling, the quality scores of all frames are classified into two groups composed of lower and higher quality using k-means clustering [42] along the temporal dimension with $k = 2$. Let $\mathbf{G}_L$ and $\mathbf{G}_H$ represent the sets of frame indices of the lower and higher quality groups, respectively. Fig. 4 illustrates the resulting clusters for a number of different video sequences. In the right column of Figure 4, green crosses correspond to frame indices in $\mathbf{G}_H$, while red dots correspond to frame indices in $\mathbf{G}_L$. The scores from the two regions are then combined to obtain an overall quality prediction for the entire video sequence:

$$S = \frac{\sum_{f \in \mathbf{G}_L} s_f + w \cdot \sum_{f \in \mathbf{G}_H} s_f}{|\mathbf{G}_L| + w \cdot |\mathbf{G}_H|} \tag{6}$$

where $|G_L|$ and $|G_H|$ denote the cardinality of G_L and G_H, respectively. The weight w is computed as a function of the ratio between the scores in $\mathbf{G_L}$ and $\mathbf{G_H}$ and is applied to the scores in the less influential higher quality region:

$$w = \left(1 - \frac{M_L}{M_H}\right)^2 \tag{7}$$

where M_H and M_L are the mean quality scores of $\mathbf{G_H}$ and $\mathbf{G_L}$, respectively. We found that while the higher quality temporal regions of the video are perceptually less influential, they cannot be completely ignored. As the difference between quality in the higher and lower quality regions increases, the influence of the higher quality regions of the video on overall quality also increases which is reflected in the definition of w. Note that when the video quality is fairly uniform and

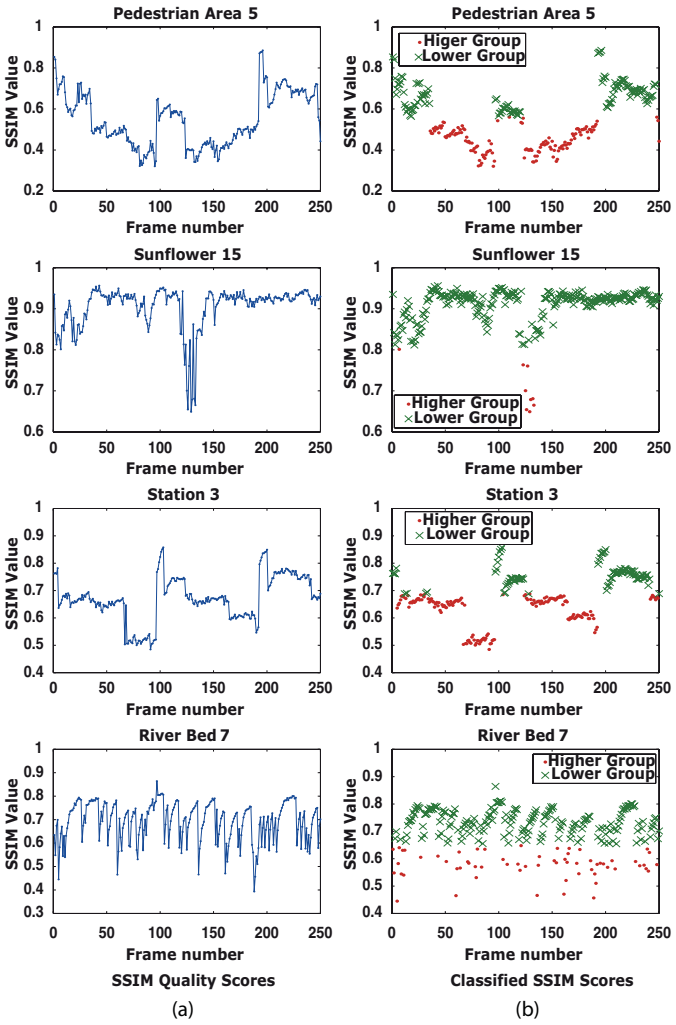

Fig. 4. Example illustrating the proposed temporal pooling method using test sequences from the LIVE Video Quality database. (a) SSIM scores computed per frame using our proposed spatial pooling method, where the x-axis denotes frame numbers. (b) Scores divided into higher and lower quality groups where crosses and dots correspond to quality scores from the higher and lower quality groups, respectively.

$M_L = M_H$, $\omega = 0$ and overall quality is determined by the lower quality frames of the video. As the difference between M_H and M_L increases, ω increases and assigns higher weights to the higher quality regions of the video in determining overall video quality.

IV. Performance

We evaluated the performance of VQPooling on the LIVE Video Quality Database and on the EPFL-PoliMI database using the MSE, SSIM and MOVIE algorithms as inputs to the VQPooling method [29], [43]. We used two databases to remove any biases in analysis or interpretation that might be incurred by only using content from one database [13].

The LIVE Video Quality Database includes 10 original reference videos and 150 distorted videos. All of the test sequences in the LIVE database are progressively scanned and have a resolution of 768×432 pixels. 15 test sequences were created from each of the reference sequences using four different distortion processes: MPEG-2 compression, H.264/AVC compression, and simulated transmission of H.264/AVC compressed bitstreams through error-prone IP networks and wireless networks. Among the 15 kinds of test sequences, four MPEG-2 compressed videos are encoded by the MPEG-2 reference software available from the International Organization for Standardization (ISO) with the compression rates varied from 700 kbps to 4Mbps [44]. Four H.264/AVC compressed videos are encoded by the JM reference software (Version 12.3) made available by the Joint Video Team (JVT) [45]. Three IP network error patterns were supplied by the Video Coding Experts Group (VCEG) [46], with loss rates of 3%, 5%, 10% and 20% and compression rates between 0.5-7 Mbps. Four wireless network error patterns were simulated by the software available from the VCEG [47], with packet error rates varied between 0.5-10% and compression rates varied between 0.5-7 Mbps. The test sequences and difference mean opinion scores (DMOS) are available in [29]. The authors of [43] provide related information, subjective VQA results on the LIVE Database and a comparison of several state-of-the-art VQA methods.

The EPFL-PoliMI database include 12 original reference sequences, half of which are 4CIF resolution (704×576 pixels) and the remaining are CIF resolution (352 × 288 pixels) [30]–[49]. All of the reference videos are encoded with the H.264/AVC reference software (Version 14.2) resulting in 12 packet loss free H.264/AVC coded bitstreams. For each of the coded 12 bitstreams, channel distortion patterns were generated at 6 different packet loss rates (PLR) (0.1%, 0.4%, 1%, 3%, 5%, 10%) and two channel realizations were selected for each PLR, resulting in 144 channel distorted bit streams. The test sequences and mean opinion scores (MOS) are available in [30]. We used the MOS scores provided from the subjective studies conducted at EPFL to report all the results in this paper.

We applied VQPooling to quality maps obtained from MSE, the SSIM index [4] and the MOVIE index [6] on the LIVE Video Quality Database. MSE is still commonly used, despite its well known perceptual shortcomings. SSIM is a popular still image QA algorithm which can be applied frame-by-frame on video. MOVIE is a state-of-the-art perception-driven VQA algorithm that operates in the spatio-temporal domain. These algorithms represent diverse approaches to VQA. However, VQPooling can be applied to the responses of any VQA algorithm that can produce local spatio-temporal estimates of video quality.

To obtain quality maps of MSE and SSIM, we utilized a sampling window of 16 × 16 that slides in increments of 4 pixels to make each measurement. In other words, MSE and SSIM are evaluated at every 4^{th} pixel along each dimension. In the case of MSE, the final overall quality index computed using different pooling methods is converted to peak signal-to-noise ratio (PSNR). This makes the results presented in Tables V and VI comparable to the results in Tables I and II. The quality map of MOVIE is obtained using the released software implementation of MOVIE [6]. There are three different versions of the MOVIE index : spatial MOVIE (SMOVIE), temporal MOVIE (TMOVIE) and MOVIE. The SMOVIE and TMOVIE primarily capture spatial and temporal

TABLE I
SROCC RESULTS ON THE LIVE VIDEO QUALITY DATABASE.
(W: WIRELESS. I: IP. H: H.264/AVC. M: MPEG2)

VQA	W	I	H	M	All
VSNR	0.7019	0.6894	0.6460	0.5915	0.6755
VQM	0.7214	0.6383	0.6520	0.7810	0.7026
PSNR	0.6574	0.4166	0.4585	0.3862	0.5397
MeanSSIM	0.5233	0.4550	0.6514	0.5545	0.5257
MOVIE	0.8019	0.7157	0.7664	0.7733	0.7890
MSE (Percentile)	0.6720	0.5715	0.5488	0.4423	0.5908
SSIM (Percentile)	0.7696	0.7428	0.7032	0.6632	0.7659
MOVIE (Percentile)	0.7992	0.7121	0.7386	0.7654	0.7650
MSE (VQPooling)	0.6958	0.5786	0.5977	0.5331	0.6470
SSIM (VQPooling)	0.8339	0.7770	0.8088	0.8275	0.8369
MOVIE (VQPooling)	0.8026	0.8060	0.8309	0.8504	0.8427

TABLE II
LCC RESULTS ON THE LIVE VIDEO QUALITY DATABASE.
(W: WIRELESS. I: IP. H: H.264/AVC. M: MPEG2)

VQA	W	I	H	M	All
VSNR	0.6992	0.7341	0.6216	0.5980	0.6896
VQM	0.7324	0.6480	0.6459	0.7860	0.7236
PSNR	0.6689	0.4645	0.5492	0.3891	0.5621
MeanSSIM	0.5401	0.5119	0.6656	0.5491	0.5444
MOVIE	0.8386	0.7622	0.7902	0.7595	0.8116
MSE (Percentile)	0.7191	0.5778	0.5780	0.4763	0.6198
SSIM (Percentile)	0.7954	0.7905	0.7339	0.6711	0.7829
MOVIE (Percentile)	0.8174	0.7631	0.7479	0.7702	0.7946
MSE (VQPooling)	0.7044	0.5383	0.6325	0.5174	0.6551
SSIM (VQPooling)	0.8526	0.8170	0.8234	0.8181	0.8511
MOVIE (VQPooling)	0.8502	0.8015	0.8444	0.8453	0.8611

distortions in the video respectively. The overall MOVIE index is defined as the product of SMOVIE and TMOVIE.

Figure 5 shows the Linear Correlation Coefficient (LCC) results of VQPooling using the LIVE database as a function of t_S on MSE, SSIM and MOVIE quality maps with $t_M = 1$ and as t_S ranges from 1 to 10 in unit increments. It can be seen that the performance of VQPooling is not very sensitive to the value of t_S for this range of values. As described in Section III-B, we chose t_S such that $t_M < t_S$ with $t_M = 1$ and $t_S = 3$. All parameters described here were identical for the results that we present on both the LIVE Video Quality Database and EPFL-PoliMI databases for all three VQA algorithms: MSE, SSIM and MOVIE.

We compare VQPooling to both mean based pooling and the percentile pooling proposed in [24]. For mean based pooling, the local spatio-temporal quality scores were pooled using the mean along both spatial and temporal dimensions. For PSNR, we computed the mean of the MSE values in the spatial and temporal dimensions before converting the overall MSE into PSNR. For percentile pooling, the scale factor r in Equation (5) weighting the lowest $p = 5\%$ was not found to influence the results significantly and we hence choose $r = 0$.

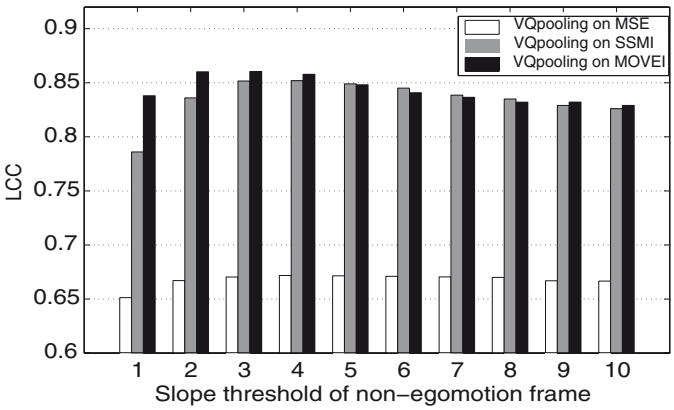

Fig. 5. LCC results of VQPooling on MSE, SSIM, and MOVIE maps as a function of t_S on the LIVE Video Quality database with $t_M = 1$.

TABLE III
SROCC RESULTS ON EPFL-POLIMI DATABASE

VQA	SROCC	VQA	SROCC
VSNR	0.8958	MSE (Percentile)	0.8808
VQM	0.8375	SSIM (Percentile)	0.9262
MSE	0.7983	MOVIE (Percentile)	0.9078
MeanSSIM	0.8357	MSE (VQPooling)	0.8821
MOVIE	0.9203	SSIM (VQPooling)	0.9471
		MOVIE (VQPooling)	0.9335

TABLE IV
LCC RESULTS ON EPFL-POLIMI DATABASE

VQA	LCC	VQA	LCC
VSNR	0.8955	MSE (Percentile)	0.8834
VQM	0.8433	SSIM (Percentile)	0.9265
MSE	0.7951	MOVIE (Percentile)	0.9184
MeanSSIM	0.8341	MSE (VQPooling)	0.850
MOVIE	0.9302	SSIM (VQPooling)	0.9543
		MOVIE (VQPooling)	0.9422

Figure 6 shows scatter plots of subjective scores and the VQPooling scores obtained using MSE, SSIM and MOVIE maps on the LIVE and EPFL-PoliMI databases. The dotted line in Fig. 6 is the best fitting logistic function of the objective scores to the subjective data. We used the logistic function specified in [50]:

$$S'_j = b_2 + \frac{b_1 - b_2}{1 + e^{-(S_j - b_3/b_4)}} \tag{8}$$

where S_j is the quality score of the j^{th} video sequence and the fitting parameters (b_1, b_2, b_3, b_4) are obtained by minimizing the least square error between the DMOS values and the fitted scores, S'_j.

Spearman Rank Order Correlation coefficient (SROCC) and Pearson LCC were used as measures to evaluate the performance of VQPooling. SROCC measures the monotonicity of objective quality scores with respect to subjective quality scores, while the LCC measures the linear accuracy of the objective quality scores. SROCC and the LCC of VQPooling relative to human subjective judgments are compared against the performance of several VQA algorithms in Table I–IV. The results show that the performance of VQPooling is highly

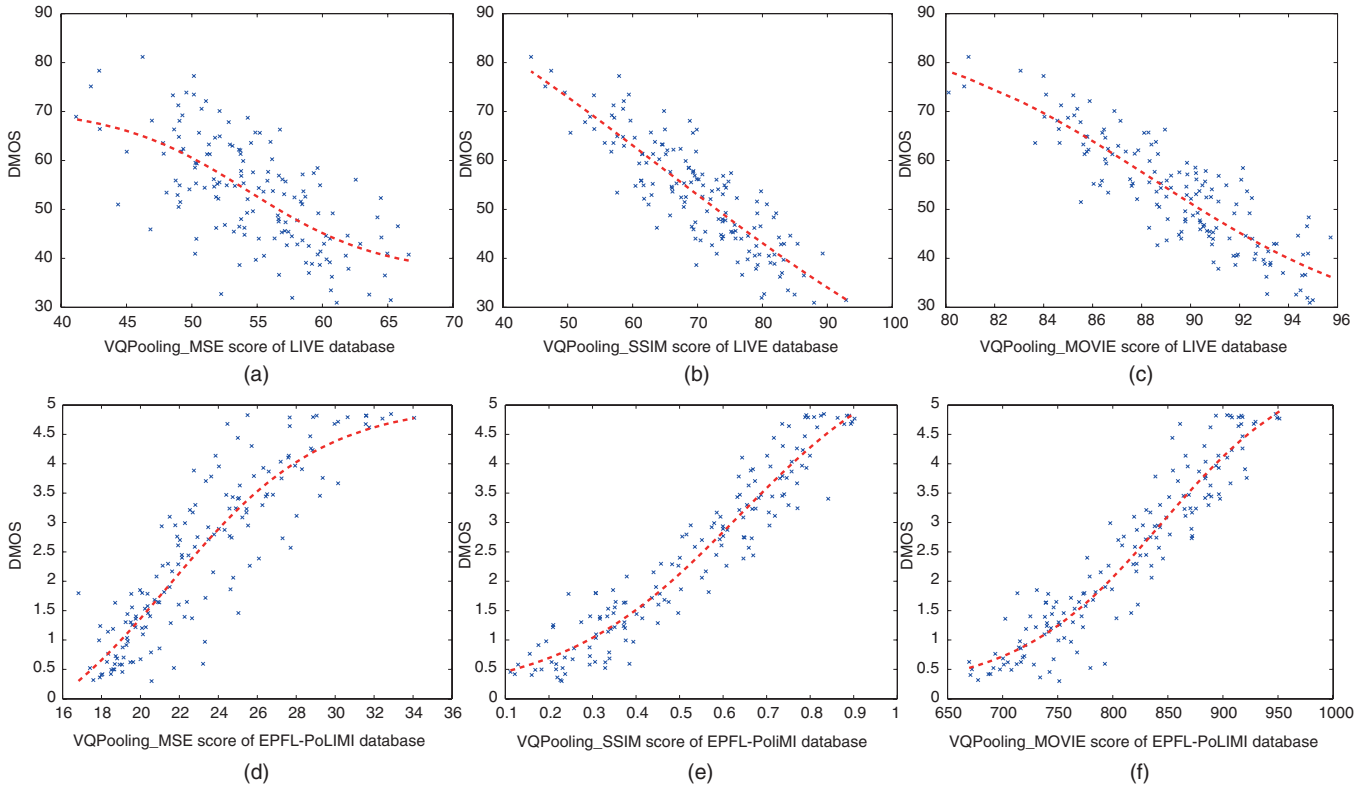

Fig. 6. Scatter plots of VQPooling scores versus DMOS values and VQPooling scores versus MOS values for all videos in the LIVE Video Quality database and the EPFL-PoLIMI database, respectively. (a) VQPooling on MSE map of LIVE database. (b) VQPooling on SSIM quality map of LIVE database. (c) VQPooling on MOVIE quality map of LIVE database. (d) VQPooling on MSE map of EPFL-PoliMI database. (e) VQPooling on SSIM quality map of EPFL-PoliMI database. (f) VQPooling on MOVIE quality map of EPFL-PoliMI database.

TABLE V
SROCC RESULTS FOR DIFFERENT POOLING STRATEGIES ON THE LIVE VIDEO QUALITY DATABASE. SCHEME A: SPATIAL MEAN AND TEMPORAL VQPOOLING. SCHEME B: SPATIAL VQPOOLING AND TEMPORAL MEAN. SCHEME C: SPATIAL AND TEMPORAL VQPOOLING

Quality Map	Scheme A	Scheme B	Scheme C
MSE	0.5642	0.6103	0.6470
SSIM	0.6011	0.7884	0.8369
MOVIE	0.6137	0.7726	0.8427

TABLE VI
LCC RESULTS FOR DIFFERENT POOLING STRATEGIES ON THE LIVE VIDEO QUALITY DATABASE. SCHEME A: SPATIAL MEAN AND TEMPORAL VQPOOLING. SCHEME B: SPATIAL VQPOOLING AND TEMPORAL MEAN. SCHEME C: SPATIAL AND TEMPORAL VQPOOLING

Quality Map	Scheme A	Scheme B	Scheme C
MSE	0.5902	0.6351	0.6551
SSIM	0.6108	0.8198	0.8511
MOVIE	0.6231	0.8023	0.8611

competitive for all distortion types. In addition, VQPooling improves the performance of all three quality indices (MSE, SSIM and MOVIE indices). The performance of VQPooling improves relative to fixed percentile pooling for all algorithms as well, showing that pooling in a content adaptive manner can improve the performance of objective VQA algorithms against human quality judgments. We also show the contribution of Spatial and Temporal VQPooling separately using the LIVE Video Quality Database in Tables V and VI.

In Tables VII and VIII, we also conducted tests for statistical significance of the results of the different pooling methods (mean, percentile, VQPooling) against each other on the LIVE Video Quality Database and the EPFL-PoliMI databases. A symbol value of "1" indicates that the statistical performance of the VQA model in the row is superior to that of the model in the column. A symbol value of "0" indicates that the statistical performance of the model in the row is inferior to that of the model in the column and "-" indicates that the statistical performance of the model in the row is equivalent to that of the model in the column. On the LIVE Video Quality Database, the SSIM index showed the most consistent improvement while MSE showed no improvement that was significant, perhaps another reason not to use MSE for VQA. The MOVIE index also showed significant improvement using VQPooling as compared to mean based and percentile pooling, which is remarkable given the already high performance of the index and its use of advanced temporal perceptual models. VQPooling improved the performance of all methods on the EPFL PoliMI database, but the improvement over percentile pooling was not found to be statistically significant.

Finally, we briefly discuss the computational complexity of VQPooling. Our un-optimized implementation of the

TABLE VII

RESULTS OF THE F-TEST PERFORMED ON THE RESIDUALS BETWEEN MODEL PREDICTIONS AND DMOS VALUES OF THE LIVE VIDEO QUALITY DATABASE FOR THE DIFFERENT POOLING STRATEGIES CONSIDERED (MEAN BASED POOLING, PERCENTILE POOLING, AND VQPOOLING RESPECTIVELY). M1–M3 CORRESPOND TO MEAN BASED POOLING, PERCENTILE POOLING AND VQPOOLING RESPECTIVELY. EACH ENTRY IN THE TABLE IS A CODEWORD CONSISTING OF 5 SYMBOLS. THE SYMBOLS CORRESPOND TO THE "WIRELESS," "IP," "H.264/AVC," "MPEG-2," AND "ALL DATA" DISTORTION CATEGORIES IN THE LIVE VIDEO QUALITY DATABASE IN THAT ORDER

	M1	M2	M3
M1	- - - - -	- - - - -	- - - - -
M2	- - - - -	- - - - -	- - - - -
M3	- - - - -	- - - - -	- - - - -

(a) PSNR

	M1	M2	M3
M1	- - - - -	0 0 - - 0	0 0 0 0 0
M2	1 1 - - 1	- - - - -	- - - - 0
M3	1 1 1 1 1	- - - - 1	- - - - -

(b) SSIM

	M1	M2	M3
M1	- - - - -	- - - - -	- - - - 0
M2	- - - - -	- - - - -	- - - - 0
M3	- - - - 1	- - - - 1	- - - - -

(c) MOVIE

TABLE VIII

RESULTS OF THE F-TEST PERFORMED ON THE RESIDUALS BETWEEN MODEL PREDICTIONS AND MOS VALUES OF THE EPFL-POLIMI DATABASE FOR THE DIFFERENT VQA ALGORITHMS. L1–L9 CORRESPOND TO PSNR, MEAN SSIM, MOVIE, MSE (PERCENTILE), SSIM (PERCENTILE), MOVIE (PERCENTILE), MSE (VQPOOLING), SSIM (VQPOOLING), AND MOVIE (VQPOOLING), RESPECTIVELY

	L1	L2	L3	L4	L5	L6	L7	L8	L9
L1	-	-	0	0	0	0	0	0	0
L2	-	-	0	0	0	0	0	0	0
L3	1	1	-	1	-	-	1	-	-
L4	1	1	0	-	0	0	-	0	0
L5	1	1	-	1	-	-	1	-	-
L6	1	1	-	1	-	-	1	0	-
L7	1	1	0	-	0	0	-	0	0
L8	1	1	-	1	-	1	1	-	-
L9	1	1	-	1	-	-	1	-	-

VQPooling algorithm, running on Matlab on a 3 GHz processor with 4 GB RAM running Windows 7, executes in about 50 seconds on the videos in the LIVE Video Quality Database (768×432 pixels, 25 fps, 10 second clips). The spatial VQPooling algorithm involve sorting the quality scores which takes 2 seconds and the temporal VQPooling algorithm involves a clustering step which takes 0.4 sec. The main computational burden is caused by estimation of the MV's for egomotion detection. We adopted the motion estimation algorithm of a practical video codec [41] in our implementation of VQPooling. Recent fast motion estimation implementations that allow for realtime processing [51], [52] can help speed up VQPooling. Further, if the VQA algorithm already utilizes motion information (for example, MOVIE), these can be reused resulting in no additional computational cost associated with MV estimation.

V. CONCLUSION

We proposed a content adaptive pooling strategy that emphasizes perceptually annoying poor quality regions in a video when predicting overall video quality. This pooling strategy is based on the observed distributions of spatio-temporally local quality scores when videos are distorted by compression or lossy transmission over a communication network. We extract influential quality scores in a content adaptive manner by studying their rank-order distribution, emphasizing annoying high distortion scores to produce a final quality score. We also studied the effects of large, egomotion-like image flow on the perception of quality and used a modified criterion to extract perceptually influential low quality regions based on the presence of such large, cohesive motion fields.

We tested our proposed VQPooling algorithm on quality maps obtained using MSE, SSIM and the MOVIE indices on the LIVE VQA Database and the EPFL-PoliMI Video Quality database [29], [30]. When VQPooling is applied to quality indices such as MSE, SSIM and MOVIE, consistent improvement in performance was observed across all distortion types and on both databases as compared to conventional pooling methods such as mean based or percentile pooling. We found that the adaptive extraction of perceptually influential low quality scores based on the distribution of scores improves the performance of mainstream competitive objective VQA algorithms.

The focus of our work has been based on distortions arising from compression and transmission over lossy communication channels which is typical in most video communication applications such as streaming video, Video on Demand, video teleconferencing and so on. Our validation of the proposed method on different databases and using different algorithms demonstrates the general applicability of this method for videos arising from such distortion sources. Further, the temporal pooling in our work has focused on analyzing short duration video segments and does not consider effects such as recency, which shows that the most recently seen part of the sequence has a heavier influence on overall perceived quality [14]. Our work also does not attempt to incorporate hysteresis effects that have been observed in human studies of video quality [17]. Study of these perceptual effects on temporal video pooling and incorporating these into VQPooling is an area of future research.

REFERENCES

[1] S. Winkler, "Issues in vision modeling for perceptual video quality assessment," *Signal Process.*, vol. 78, pp. 231–252, Oct. 1999.

[2] A. C. Bovik, *The Handbook of Image and Video Processing*. New York: Academic, 2005.

[3] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[5] M. 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.

[6] 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.

[7] 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.

[8] Z. Wang, L. Lu, and A. C. Bovik "Video quality assessment based on structual distortion measurement," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 121–132, Feb. 2004.

[9] Z. Wang and Q. Li, "Video quality assessment using a statistical model of human visual speed perception," *J. Opt. Soc. Amer. A*, vol. 24, no. 12, pp. B61–B69, Dec. 2007.

[10] C. J. B. Lambrecht and O. Verscheure, "Perceptual quality measure using a spatio-temporal model of the human visual system," *Proc. SPIE*, vol. 2668, pp. 450–461, Mar. 1996.

[11] A. Ninassi, O. L. Meur, P. L. 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.

[12] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "On the performance of human visual system based image quality assessment metric using wavelet domain," *Proc. SPIE Human Vis. Electron. Imag. XIII*, vol. 6806, pp. 680610-1–680610-12, Dec. 2008.

[13] F. M. Ciaramello and S. S. Hemami, "The influence of space and time varying distortions on objective intelligibility estimators for region-of-interest video," in *Proc. IEEE Int. Conf. Image Process.*, Hong Kong, Sep. 2010, pp. 1097–1100.

[14] D. E. Pearson, "Viewer response to time-varying video quality," in *Human Vision and Electronic Imaging III*, vol. 3299, B. E. Rogowitz and T. N. Pappas, Eds. Bellingham, WA: SPIE, 1998, pp. 16–25.

[15] M. Barkowsky, B. Eskofier, R. Bitto, J. Bialkowski, and A. Kaup, "Perceptually motivated spatial and temporal integration of pixel based video quality measures," in *Welcome to Mobile Content Quality of Experience*. Vancouver, BC, Canada: ACM, 2007, pp. 1–7.

[16] M. A. Masry and S. S. Hemami, "A metric for continuous quality evaluation of compressed video with severe distortions," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 133–146, Feb. 2004.

[17] K. Seshadrinathan and A. C. Bovik, "Temporal hysteresis model of time varying subjective video quality," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2011, pp. 1153–1156.

[18] F. Yang, S. Wan, Q. Xie, and H. R. Wu, "No-reference quality assessment for networked video via primary analysis of bit stream," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1544–1554, Nov. 2010.

[19] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video quality assessment," *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 129–132, Mar. 2002.

[20] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 977–992, Jul. 2001.

[21] S. Lee and A. C. Bovik, "Fast algorithms for foveated video processing," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 13, no. 2, pp. 149–162, Feb. 2003.

[22] S. Lee, A. C. Bovik, and Y. Y. Kim, "High quality, low delay foveated visual communications over mobile channels," *J. Visual Commun. Image Represent.*, vol. 16, no. 2, pp. 180–211, Apr. 2005.

[23] J. You, J. Korhonen, and A. Perkis, "Attention modeling for video quality assessment: Balancing global quality and local quality," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2010, pp. 914–919.

[24] A. K. Moorthy and A. C. Bovik "Visual importance pooling for image quality assessment," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 193–201, Apr. 2009.

[25] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Proc. IEEE Int. Conf. Image Process.*, Atlanta, GA, Oct. 2006, pp. 2945–2948.

[26] S. Rimac-Drlje, M. Vranjes, and D. Zagar, "Influence of temporal pooling methdod on the objective video quality evaluation," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, Jul. 2009, pp. 1–5.

[27] N. Suresh and N. Jayant, "'Mean time between failures': A subjectively meaningful video quality metric," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2006, pp. 941–944.

[28] M. Narwaria, W. Lin, and A. Liu, "Low-complexity video quality assessment using temporal quality variations," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 525–535, Jun. 2012.

[29] *LIVE Video Quality Database*. (2009) [Online]. Available: http://live.ece.utexas.edu/research/quality/livevideo.html

[30] *EPFL-PoliMI Video Quality Assessment Database*. (2009) [Online]. Available: http://vqa.como.polimi.it/

[31] M. Yuen and H. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," *Signal Process.*, vol. 70, no. 3, pp. 247–278, 1998.

[32] J. Park, K. Seshadrinathan, S. Lee, and A. C. Bovik, "Spatio temporal quality pooling accounting for transient severe impairment and egomotion," in *Proc. IEEE Int. Conf. Image Process.*, Brussels, Belgium, Sep. 2011, pp. 2509–2512.

[33] A. C. Bovik, *The Essential Guide to Image Processing*. New York: Academic, 2009.

[34] S. Lee, M. S. Pattichis, and A. C. Bovik, "Foveated video compression with optimal rate control," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 977–992, Jul. 2001.

[35] J. Park, H. Lee, S. Lee, and A. C. Bovik, "Optimal channel adaptation of scalable video over a multicarrier-based multicell environment," *IEEE Trans. Multimedia*, vol. 11, no. 6, pp. 1062–1071, Oct. 2009.

[36] U. Jang, H. Lee, and S. Lee, "Optimal carrier loading control for the enhancement of visual quality over OFDMA cellular networks," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1181–1196, Oct. 2008.

[37] X. Yi and N. Ling, "Improved H.264 rate control by enhanced MAD-based frame complexity prediction," *J. Visual Commun. Image Represent.*, vol. 17, pp. 407–424, Dec. 2006.

[38] M. Jiang and N. Ling, "Low-delay rate control for real-time H.264/AVC video coding," *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 467–477, Jun. 2006.

[39] A. A. Stocker and E. P. Simoncelli, "Noise characteristics and prior expectations in human visual speed perception," *Nature Neurosci.*, vol. 9, pp. 578–585, Mar. 2006.

[40] J. W. Suchow and G. A. Alvarez, "Motion silences awareness of visual change," *Current Biol.*, vol. 21, pp. 1–4, Jan. 2011.

[41] *ITU Recommendation H.263: Video Coding for Low Bit Rate Communication*. (2001) [Online]. Available: http://www.itu.int/rec/T-REC-H.263/en

[42] J. A. Hartigan, *Clustering Algorithms*. New York: Wiley, 1975.

[43] 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.

[44] *Generic Coding of Moving Pictures and Associated Audio Information. Part 2: Video*. (2004, Nov.) [Online]. Available: http://standards.iso.org/ittf/PubliclyAvailable-Standards/c039486ISOIEC13818-52005ReferenceSoftware.zip

[45] *H.264/AVC Software Coordination*. (2007) [Online]. Available: http://iphome.hhi.de/suehring/tml/

[46] *Proposed Error Patterns for Internet Experiments*. (1999) [Online]. Available: http://ftp3.itu.ch/av-arch/video-site/9910_Red/q15i16.zip

[47] *Common Test Conditions for RTP/IP Over 3GPP/3GPP2*. (1999) [Online]. Available: http://ftp3.itu.ch/av-arch/video-site/0109_San/VCEG-N80_software.zip

[48] F. D. 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. Qual. Multimedia Experience*, San Diego, CA, Jul. 2006, pp. 204–209.

[49] F. D. Simone, M. Naccari, F. Dufaux, M. Tagliasacchi, S. Tubaro, and T. Ebrahimi, "H.264/AVC video database for the evaluation of qualitiy metrics," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2010, pp. 2430–2433.

[50] *Validation of Objective Quality Metrics for Video Quality Assessment*. (2000) [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtvphaseI

[51] M. Kim, I. Hwang, and S.-I. Chae, "A fast vlsi architecture for fullsearch variable block size motion estimation in MPEG-4 AVC/H.264," in *Proc. Asia South Pacific Design Autom. Conf.*, vol. 1. Jan. 2005, pp. 631–634.

[52] R. Strzodka and C. Garbe, "Real-time motion estimation and visualization on graphics cards," in *Proc. IEEE Visual. Conf.*, Mar. 2004, pp. 545–552.

Jincheol Park was born in Korea in 1982. He received the B.S. degree in information and electronic engineering from Soongsil University, Seoul, Korea, in 2006, and the M.S. degree in electrical and electronic engineering from Yonsei University, Seoul, in 2008, where he is currently pursuing the Ph.D. degree with the Multidimensional Insight Laboratory.

He was a Visiting Researcher under the guidance of Prof. A. C. Bovik with the Laboratory for Image and Video Engineering, Department of Electrical and Computer Engineering, University of Texas, Austin, from June 2010 to June 2011. His current research interests include 2-D and 3-D video quality assessment.

Kalpana Seshadrinathan (S'03–M'09) received the B.Tech. degree in technology from the University of Kerala, Thiruvananthapuram, India, in 2002, and the M.S. and Ph.D. degrees in electrical engineering from the University of Texas at Austin in 2004 and 2008, respectively.

She is currently a Research Scientist with Intel Corporation, Santa Clara, CA. Her current research interests include image and video quality assessment, computational aspects of human vision, mobile imaging, computational photography, and statistical modeling of images and video.

Dr. Seshadrinathan was the recipient of the Texas Telecommunications Engineering Consortium Graduate Fellowship in 2003 and the Graduate Student Professional Development Award from the University of Texas at Austin in 2007. She was the General Co-Chair of the International Workshop on Video Processing and Quality Metrics in 2012 and the Assistant Director of the Laboratory for Image and Video Engineering at the University of Texas at Austin from 2005 to 2008.

Sanghoon Lee (M'05–SM'12) was born in Korea in 1966. He received the B.S. degree from Yonsei University, Seoul, Korea, and the M.S. degree from the Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 1989 and 1991, respectively, and the Ph.D. degree from the University of Texas at Austin in 2000, all in electrical engineering.

He was with Korea Telecom, Seoul, from 1991 to 1996. In 1999, he was with Bell Laboratories, Lucent Technologies, New Providence, NJ, and was involved in research on wireless multimedia communications. From 2000 to 2002, he was involved in development of real-time embedded software and communication protocols for 3G wireless networks with Lucent Technologies. In 2003, he joined the Department of Electrical and Electronics Engineering, Yonsei University, where he is currently a Full Professor. His current research interests include image and video quality assessments, image and video processing, wireless multimedia communications, and 4G wireless networks.

Dr. Lee was a recipient of the Special Service Award from the IEEE Broadcast Technology Society in 2012. He is an Associate Editor of the IEEE TRANSACTION ON IMAGE PROCESSING and an Editor of the *Journal of Communications and Networks*, the Chair of the IEEE Standard Working Group for 3-D Quality Assessment, a Guest Editor of the IEEE TRANSACTION ON IMAGE PROCESSING in 2012, and the General Chair of the IEEE IVMSP Workshop in 2013.

Alan Conrad Bovik (S'80–M'81–SM'89–F'96) is currently the Curry/Cullen Trust Endowed Chair Professor with the University of Texas at Austin, Austin, where he is the Director of the Laboratory for Image and Video Engineering and a Faculty Member with the Department of Electrical and Computer Engineering and the Center for Perceptual Systems, Institute for Neuroscience. His current research interests include image and video processing, computational vision, and visual perception. He has authored or co-authored more than 650 technical articles in these areas and holds two U.S. patents. His several books include the recent companion volumes *The Essential Guides to Image and Video Processing* (Academic Press, Waltham, MA, 2009).

Dr. Bovik was a recipient a number of major awards from the IEEE Signal Processing Society, including the Meritorious Service Award in 1998, the Technical Achievement Award in 2005, the Best Paper Award in 2009, and the Education Award in 2007. He was named the SPIE/IS&T Imaging Scientist of the Year in 2011. He was a recipient of the Hocott Award for Distinguished Engineering Research from the University of Texas at Austin in 2008, the Distinguished Alumni Award from the University of Illinois at 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 was on the Board of Governors, IEEE Signal Processing Society, from 1996 to 1998, the Editor-in-Chief of IEEE TRANSACTIONS ON IMAGE PROCESSING from 1996 to 2002, on the Editorial Board of the PROCEEDINGS OF THE IEEE from 1998 to 2004, 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 the state of Texas and is a frequent consultant to legal, industrial, and academic institutions.