Efficient Video Quality Assessment Along Temporal Trajectories

Anush Krishna Moorthy and Alan Conrad Bovik, *Fellow, IEEE*

***Abstract*—We propose a new video quality assessment (VQA) algorithm—the motion compensated structural similarity index—that assesses not only spatial quality but also quality along temporal trajectories. Drawing inspiration from the motion-compensated approach followed for video compression, we propose a motion-compensated approach to temporal quality assessment. The proposed algorithm is computationally efficient as compared to other VQA algorithms that utilize motion information from extracted optical flow and correlates well with human perception of quality. In order to exemplify the utility of the algorithm in a practical setting, we evaluate the quality of H.264/AVC compressed videos. Efficiency of computation is enabled by the novel motion-vector re-use concept.**

***Index Terms*—H.264/AVC, motion-compensation, spatio-temporal quality assessment, structural similarity, video quality assessment.**

I. Introduction

Objective video quality assessment (VQA) refers to evaluation of the quality of videos by an algorithm where the goal is to produce scores which correlate well with human perception of quality. Perceived quality is generally gauged by conducting a large scale subjective study where a group of human observers are asked to rate a set of videos on a particular scale, and their scores are averaged to form a mean opinion score (MOS). In this letter, our focus will be on full reference (FR) objective quality assessment algorithms, where the algorithm has access to both the reference and distorted videos.

Distortions in videos can occur spatially and temporally.[1] Spatial distortions include blocking, blurring, ringing, mosaic patterns, and so on, while temporal distortions include motion compensation mismatches, mosquito effects, ghosting, smearing, and so on. A review of different kinds of distortions that occur in compressed videos can be found in [1]. Simply listing the various kinds of distortion in videos suggests that a spatial-only quality metric will fail to capture many perceptually relevant temporal distortions [2]–[5]. Here, we develop an algorithm that assesses both spatial and temporal quality and demonstrate that such an approach produces noticeable improvements in terms of correlation with human perception.

Although mean squared error and peak signal-to-noise ratio (PSNR) have been extensively applied for VQA, many researchers in the past have pointed out that it correlates poorly with human perception of quality [6], [7]. One approach to developing VQA algorithms is by attempting to model human visual system mechanisms [8]–[12] . Although a human visual system (HVS) based system seems like an ideal route to take, much work is left to be done in understanding human visual processing [13]. Until research in vision science allows for a complete and precise modeling of the HVS, measures of quality based on the HVS are likely to fall short of accurately predicting quality of videos.

In [14], the authors proposed a simple idea for VQA, by extending the image quality assessment measure—single-scale structural similarity index (SS-SSIM) [14], where SS-SSIM was applied on a frame-by-frame basis. The authors also proposed the use of a weighting scheme that took into account some motion estimated using a block-motion estimation algorithm. In [3], the authors used an alternate weighting scheme based on human perception of motion information. In both these cases, spatial quality computed using SS-SSIM was weighted based on motion information. However, temporal-based weighting of spatial quality scores does not necessarily account for temporal distortions [2]. As mentioned before, temporal distortions can differ significantly from spatial distortions. Further, vision research has hypothesized that the HVS has (approximately) separate channels for spatial and temporal processing [15]–[17]. The weighted pooling of spatial scores does not capture this separability.

Recently, researchers in the area of VQA have started exploring the space of temporal distortions and its effects on quality [4], [5], even though they choose not to test their algorithms on a public dataset. In [4], the authors proposed multi-stage model that incorporates spatio-temporal "tubes" over average fixation durations and a host of thresholds and pooling techniques. The algorithm in [5] again utilizes motion information in conjunction with attentional modeling, thresholds and varied pooling strategies. The authors also model frame-rate pauses and skips to estimate video quality. In [2], the authors utilized properties of the neurons in the visual cortex including spatial frequency and orientation selectivity to develop the motion-based video integrity evaluation (MOVIE).

Manuscript received September 4, 2009; revised January 7, 2010 and April 20, 2010; accepted July 17, 2010. Date of publication October 14, 2010; date of current version November 5, 2010. This paper was recommended by Associate Editor S.-Y. Chien.

The authors are with the Laboratory for Image and Video Engineering (LIVE), Department of Electrical and Computer Engineering, University of Texas, Austin, TX 78712 USA (e-mail: anushmoorthy@mail.utexas.edu; bovik@ece.utexas.edu).

Digital Object Identifier 10.1109/TCSVT.2010.2087470

[1]In general, many of these distortions may be classified as spatio-temporal, but we do away with such classification here.

However, the computational complexity of the algorithm makes practical implementation difficult as it relies on 3-D optical flow computation.

In this letter, we propose an algorithm that attempts to bridge the gap between a model that represents the HVS processing accurately, and one that performs time-efficient computation of objective video quality. We describe the algorithm in detail, and evaluate the performance of the proposed algorithm on a publicly available dataset and demonstrate that it correlates well with human perception of quality. Further, in order to demonstrate a practical implementation of the algorithm, we explain how the algorithm can be used for quality assessment of compressed videos with very low overhead.

II. Motion Compensated SSIM

The strategy we will take toward adapting the SSIM concept to "true motion" VQA is to utilize SS-SSIM to evaluate spatial quality and to develop a simple temporal metric which, coupled with SSIM leads to a VQA algorithm that can be practically implemented. Temporal quality assessment in our algorithm is carried out using a combination of a block-based motion estimation algorithm and SS-SSIM. The reader will appreciate the fact that both of these operations require low computational complexity in comparison with an algorithm that relies on optical flow computation. Further, instead of using a filter-bank that requires a large temporal support, we utilize only neighboring frames in the video. Even though the critical time for temporal vision is thought to be around 200–300 ms, in a real-time scenario, our scheme will help in assessing instantaneous quality without the need to wait for enough frames to fill the buffer. This new algorithm is called the motion-compensated structural similarity index (MC-SSIM), since it evaluates structural retention between motion-compensated regions in a frame. The algorithm is explained in detail in this section.

Consider the reference video R and distorted video D of dimensions $P \times Q$ with N frames. Spatial quality computation is undertaken using SS-SSIM on a frame-by-frame basis. For each frame, the spatial quality map so obtained is of dimension (P, Q), and the spatial quality scores are denoted as $S(x, y, t), (x = \{1 \dots P\}, y = \{1 \dots Q\}, t = \{1 \dots N\})$.

Temporal quality computation proceeds as follows. In order to estimate motion, we apply a block-based motion estimation algorithm. The algorithm is applied on a frame-by-frame basis, where motion vectors are obtained for frame i from its preceding frame $i - 1$. We seek to characterize the distortion in D, and hence motion estimation is performed only on the reference video. This strategy was previously explored in [2] and [3]. In our current implementation we use adaptive-rood-pattern-search (ARPS) [18] for motion estimation. The block size is set at $b \times b$.

In order to evaluate quality, we proceed as follows. For a frame i and for block (m_R, n_R) $(m_R = \{1, 2 \dots P/b\}$, $n_R = \{1, 2 \dots Q/b\})$, in video R, we compute the motion-compensated block (m'_R, n'_R) in frame $i - 1$ by displacing the (m_R, n_R)th block by an amount indicated by the motion vector. A similar computation is performed for the corresponding (m_D, n_D)th block in D, thus obtaining the motion-compensated block (m'_D, n'_D). We then perform a quality computation between the blocks $B_R = (m'_R, n'_R)$ and $B_D = (m'_D, n'_D)$. This quality computation can be performed using any suitable image quality index. In our implementation we use SS-SSIM. Hence, for each block we obtain a quality index corresponding to the perceived quality of that block, and for each frame we obtain a quality map of dimension $(P/b, Q/b)$. We denote the temporal quality map thus obtained as $T(x, y, t), (x = \{1 \dots P/b\}, y = \{1 \dots Q/b\}, t = \{1 \dots N - 1\})$.

The original SS-SSIM proposed for image quality assessment used the mean of the local quality scores to form a single score for the image. When applied on a frame-by-frame basis on a video, the score for the video was defined as the mean value of the scores obtained from each of the frames. Alternative pooling techniques for image quality have been recently explored [19], [20]. It has been argued that the simple mean does not effectively capture the overall quality of the image. Our algorithm employs the percentile pooling strategy proposed in [21] for spatial and temporal quality to produce $S(t)$ and $T(t)$ for each frame t respectively. We note that this method is similar to the approach proposed in [22]. Alternative pooling strategies for temporal pooling remain relatively unexplored [4]. The frame-level scores are averaged to produce S and T for the spatial and temporal quality estimates respectively. The final quality score for the video is $S \times T$. As mentioned in the introduction, the spatio-temporal separability of the human visual system is reflected in the above calculation.

In our original implementation of MC-SSIM, the block size chosen for motion estimation was set at $b = 8$. However, we also tested the performance of MC-SSIM when using various other block-sizes $b = 4, b = 16$. The results for each of the chosen block-sizes are given in the next section. We find that the quality index is relatively robust to the block size.

In [23], the authors noted that the second-scale of a multi-scale image representation seems to perform extremely well in terms of correlation with human perception. The importance of the second-scale was again noted in [20], where only the second scale of the multi-scale decomposition was used to pool scores in implementing the multi-scale SSIM. As in the original implementation of SS-SSIM for video [24], we use the second scale for quality computation. Specifically, each frame is low pass filtered using a rectangular filter and then subsampled by a factor of two before quality computation.

Temporal quality is assessed not only on the "Y" component, but also on the color channels "Cb" and "Cr." For each of the channels, motion estimation is performed to extract corresponding motion vectors and the algorithm as described in the previous section is applied. Even though the results reported here utilize motion estimates from color channels, we have found that using (compensated) motion estimates from the luma channel does not affect performance. This is further demonstrated in Section V. The final temporal quality score for the video is computed as

$$T^{final} = 0.8 \times T^{Y} + 0.1 \times T^{Cb} + 0.1 \times T^{Cr}$$

where T^Y, T^{Cb}, and T^{Cr} are the temporal quality scores on each of the three color channels obtained as described above. A similar quality computation is undertaken for each of the three channels to assess spatial quality as well. The final spatial quality is computed as

$$S^{final} = 0.8 \times S^Y + 0.1 \times S^{Cb} + 0.1 \times S^{Cr}$$

where S^Y, S^{Cb}, and S^{Cr} are the spatial quality scores on each of the three color channels obtained as described above. The weights assigned to each of the channels are exactly as in [24] and are re-used here, though incorporating color in VQA remains an interesting avenue of research.

The essence of the proposed algorithm is SS-SSIM. It can easily be shown that the computational complexity of SS-SSIM is $O(PQ)$. Since we use percentile pooling there is a need to sort the SSIM scores and this can be performed with a worst-case complexity of $O(PQlog(PQ))$. The motion estimation algorithm that we use is ARPS [18]. The authors in [18] claimed that the algorithm is highly computationally efficient. They stated that ARPS performs 1.9 to 3.4 times faster than the previously proposed diamond search [25]. ARPS does not overly concern us however. It is clear that any motion-estimation algorithm could have been used for MC-SSIM. The major bottleneck in MC-SSIM is this motion-estimation phase. However, as we shall show in Section IV, we can completely avoid this bottleneck by re-utilizing motion vectors computed for compressed videos. In this case, the complexity of MC-SSIM is not much greater than that for SS-SSIM. Compare this with that of video quality metric (VQM) [22] – $O((PQ)^2)$. Further, as shown in [26], the SSIM index can be simplified without sacrificing performance. Finally, as we shall see in the next section, MC-SSIM correlates better with the human perception of quality than SS-SSIM thus making MC-SSIM an attractive VQA algorithm.

III. Results

We test MC-SSIM on the video quality experts group (VQEG) FR-TV Phase I database [27] dataset and the recently released LIVE VQA database [28]. The VQEG dataset consists of 20 reference and associated 320 distorted videos with the differential mean opinion score (DMOS) for each distorted video. The LIVE VQA dataset consists of 10 reference videos with 15 distortions each, to give a total of 150 distorted videos with the associated DMOS. The LIVE VQA dataset contains a wide range of distortions including compression and packet loss and was created to overcome the drawbacks associated with the VQEG set.

In accordance with VQEG recommendations, the evaluation measures used are the Spearman rank ordered correlation coefficient (SROCC), the Pearson (linear) correlation coefficient (LCC) and outlier ratio (OR). The results are shown in Table I for the VQEG dataset. The LCC was computed after fitting the scores produced by the algorithm to the DMOS using the logistic function prescribed in [27].

Table I also lists performance of various other algorithms. PSNR is a baseline for performance evaluation of any quality

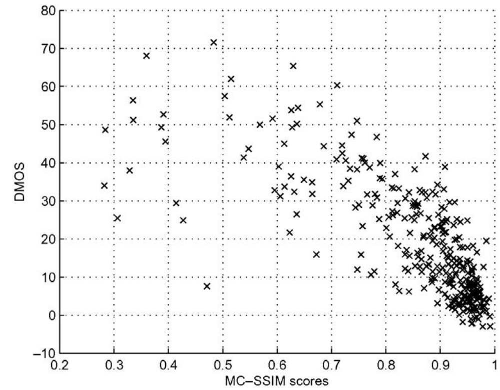

Fig. 1. Scatter plot of DMOS versus MC-SSIM scores.

TABLE I
MC-SSIM Performance (VQEG): Spearman Rank Ordered Correlation Coefficient (SROCC), Linear Correlation Coefficient (LCC), and Outlier Ratio (OR)

Algorithm	SROCC	LCC	OR
PSNR	0.782	0.779	0.678
Proponent P8 (Swisscom) [27]	0.803	0.827	0.578
SS-SSIM (no weighting) [24]	0.788	0.820	0.597
SS-SSIM (weighted) [24]	0.812	0.849	0.578
SW-SSIM (dense Y only)	0.837	0.810	0.622
MOVIE (Y only) [2]	0.833	0.821	0.644
MC-SSIM (8×8)	0.848	0.853	0.597
MC-SSIM (4×4)	0.846	0.851	0.606
MC-SSIM (16×16)	0.833	0.833	0.616

assessment algorithm. Proponent P8 was the best performing model from the ten algorithms tested by the VQEG. SS-SSIM refers to application of SS-SSIM on a frame-by-frame basis [24]. Note that this corresponds to the “no-weighting” case. We also list the performance of SS-SSIM with weighting, indicated by SS-SSIM (weighting). Another recently proposed algorithm [3], which we label speed-weighted SSIM (SW-SSIM) is listed as well. Finally, we list the performance of the recent MOVIE [2].

MC-SSIM performs competitively with these popular VQA algorithms. We also note that MC-SSIM has no “tune-able” parameters and hence its performance is reflective of its generalization capabilities.

In Fig. 1, we plot the scatter plot of MC-SSIM scores versus DMOS. From the plot, one would conjecture that MC-SSIM performs well for the high-quality case, and seems to present lowered performance for the low-quality one.

As mentioned in the previous section, we list the performance of MC-SSIM for different block sizes -4×4 and 16×16. The performance does not seem to differ much with change in block size, suggesting that the performance of MC-SSIM is robust with respect to block size.

In order to demonstrate that the use of temporal information does indeed improve the performance of the algorithm, and that the pooling strategy applied to the spatial scores is alone not responsible for the improvement, in Table II, we tabulate the SROCC, LCC and OR values for the entire VQEG database. The values show the SROCC for the spatial only

TABLE II
EFFECT OF TEMPORAL QUALITY ASSESSMENT

	SROCC	LCC	OR
SS-SSIM (no weighting) [24]	0.788	0.820	0.597
Spatial only	0.824	0.837	0.609
Spatio-temporal	0.848	0.853	0.597

TABLE III
MC-SSIM PERFORMANCE ON NATURAL VIDEO

	SROCC	LCC
MOVIE (Y only) [2]	0.860	0.858
MC-SSIM (8×8)	0.869	0.880

case (using percentile pooling) and for the full spatio-temporal implementation (as described before). Obvious improvements are seen. We also tabulate SS-SSIM scores from [24]. The pooling technique does improve the performance of SS-SSIM however, coupling the spatial score with the temporal score boosts the performance even higher.

Even though MC-SSIM does not seek to explicitly model the HVS, it is based on SSIM, and it was shown in [29] that SSIM relates to the NSS model for quality proposed in [30]. The statistics of natural scenes differ significantly from those for artificial scenes. The VQEG dataset consists of four artificial sequences (src4, src6, src16, src17), including scrolling text. In these cases, judging quality through an algorithm which has been developed for VQA may not be completely fair. In order to demonstrate that elimination of these artificial scenes affects the performance of MC-SSIM, in Table III we list the performance of MC-SSIM using 8×8 blocks on the entire VQEG dataset with the non-natural sequences removed from the analysis. For a comparison, we also include the results from MOVIE (Y component) [2], when applied to only natural videos. Again, the performance of MC-SSIM is highly competitive.

Table IV shows the performance of MC-SSIM on the LIVE VQA dataset with SROCC and LCC as measures of performance. Also listed are PSNR, SS-SSIM, SW-SSIM, VQM [22], and MOVIE. Note that MC-SSIM performs much better than PSNR, SS-SSIM and SW-SSIM and is competitive with VQM; however, MOVIE performs much better. As we have mentioned before, the goal was to create an efficient algorithm rather than one that beats all present-day algorithms. Having said that, a statistical significance analysis using the F-statistic on the residuals between the algorithmically predicted scores (after curve-fitting) and DMOS [31] indicates that the performance of MC-SSIM is *statistically indistinguishable* from that of VQM and SW-SSIM at the 95% confidence level and MOVIE is statistically better than MC-SSIM. A similar analysis shows that MC-SSIM performance is statistically better than PSNR and SS-SSIM.

IV. APPLICATION: QUALITY ASSESSMENT OF COMPRESSED VIDEOS

We consider an application of assessing quality of compressed videos. Since MC-SSIM is a full reference algorithm,

TABLE IV
MC-SSIM PERFORMANCE (LIVE)

Algorithm	SROCC	LCC
PSNR	0.3684	0.4035
SS-SSIM (no weighting) [24]	0.5257	0.5444
SW-SSIM [3]	0.5849	0.5962
VQM [22]	0.7026	0.7236
MOVIE [2]	0.7861	0.8102
MC-SSIM (8×8)	0.6791	0.6976

we assume that we have the pristine reference video for quality assessment. However, in this case the video available (reference video) is compressed using a suitable video compression algorithm. Such a scenario is much more practical as compared to quality assessment using an uncompressed source since users are unlikely to have the uncompressed version to test quality. Further, we assume that this compressed video passes through the proverbial "black box" and the goal is to assess the quality of the video at the output (test video) of this black box. One approach to such quality assessment is to perform a decompression of both the reference and test videos and then apply a VQA algorithm. If we choose to apply an algorithm that utilizes motion information, we would need to perform optical flow computation after decompression. It is at this point that using MC-SSIM provides a tremendous benefit in terms of computational complexity as well as performance. We note that the demonstration here is for the situation where the compressed video is transmitted over a channel and motion vectors from the original reference stream can be used for quality assessment of the received distorted video.

Most video compression standards utilize a motion-compensated frame differencing approach to compression [32], where motion vectors are computed at the encoder for compression and are extracted from the compressed stream at the decoder for reconstruction. Since MC-SSIM performs a computation mimicking the decompression process, the easiest solution to VQA using MC-SSIM is to re-use the motion vectors computed by the compression algorithm. Specifically, the motion vectors that we use for motion-compensated quality assessment will be the same as those used by the algorithm for motion compensated decompression. By re-utilizing motion vectors from the compression process, we have effectively eliminated a major bottleneck for VQA algorithms—that of computation of motion. This coupled with the fact that we use the simple SSIM for quality assessment will reduce overhead, and will allow for practical deployment of the algorithm. It is clear that any video compression algorithm that follows a motion-compensated-frame-differencing structure may be used for this purpose. In this letter, we choose to use the H.264/AVC compression standard [33].

In our implementation, we allow only for motion compensation using one previously decoded frame. Using motion estimates from multiple reference frames may be accomplished in a manner similar to that in [4]; and this will be investigated in the future. We fix the block-size for motion estimates at 16×16 and do not allow sub-pixel motion estimates. One reason for this is to allow for an objective comparison between

TABLE V
MC-SSIM PERFORMANCE USING H.264 MOTION VECTORS ON NATURAL VQEG VIDEOS

Algorithm	SROCC	LCC	OR
MC-SSIM (16×16)	0.867	0.879	0.586
H.264-MC-SSIM	0.872	0.879	0.606

MC-SSIM proposed in the earlier section and MC-SSIM applied on H.264/AVC compressed videos. Second, this "coarse" approximation will give us an approximate lower bound on the performance, since improved motion estimates and variable block sizes will only strengthen algorithm performance. Further, only the first frame is encoded as an I-frame and all other frames are P-frames. Other group-of-pictures settings remain interesting avenues for future exploration. We set the quantization parameter to 16, other QP settings remain interesting avenues of interest. The JM reference encoder is used in order to perform H.264 compression/decompression [34].

At this stage, we are in the situation described at the beginning of this section. We have with us a set of compressed videos (which we created artificially for the purpose of evaluation here) and we have a "black-box." We also have the (decompressed) videos at the output of this black-box (distorted videos from the VQEG dataset). So, all that remains to be done is decompress the compressed originals and perform quality assessment on corresponding input-output video pairs. The only addition here, as we described before is the extraction of motion vectors from the original video. Specifically, as we decompress the original video prior to quality computation, we also extract and save corresponding motion vectors from the decompression algorithm.

After having extracted motion vectors from the H.264 compressed videos, MC-SSIM is applied as described before on the decompressed reference and test videos. For the chroma channels, we follow the recommendations of the H.264 standard, where the chroma motion vectors are extracted by multiplying the luma motion vectors by a factor of 2 [33]. We use the VQEG database described before [27] as a test-bed for evaluating performance.

The results of using MC-SSIM using H.264 motion vectors on the entire VQEG database (natural sequences only) are shown in Table V, where we also list MC-SSIM using 16×16 blocks for motion estimation for a comparison. The algorithm performance is evaluated in terms of the above mentioned measures—SROCC, LCC, and OR.

V. CONCLUSION

In this letter, we proposed a new video quality assessment index—MC-SSIM. We explained the motivation behind the algorithm and demonstrated how the algorithm was implemented. The performance of the algorithm was evaluated on two publicly available databases and was compared with state-of-the-art techniques for video quality assessment. MC-SSIM was shown to perform well in terms of correlation with human perception. As an application, MC-SSIM was used for the quality assessment of compressed videos. The simplicity of the algorithm along with its extremely competitive performance makes MC-SSIM an attractive choice for practical deployment in order to perform video quality assessment.

REFERENCES

[1] 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.

[2] 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.

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

[4] 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. Top. Signal Process.*, vol. 3, no. 2, pp. 253–265, Apr. 2009.

[5] M. Barkowsky, B. E. J. Bialkowski, R. Bitto, and A. Kaup, "Temporal trajectory aware video quality measure," *IEEE J. Sel. Top. Signal Process.*, vol. 3, no. 2, pp. 266–279, Apr. 2009.

[6] B. Girod, "What's wrong with mean-squared error?" in *Digital Images and Human Vision*, A. B. Watson, Ed. Cambridge, MA: MIT Press, 1993, pp. 207–220.

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

[8] C. Van den Branden Lambrecht and O. Verscheure, "Perceptual quality measure using a spatiotemporal model of the human visual system," in *Proc. SPIE*, 1996, pp. 450–461.

[9] M. Masry, 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.

[10] S. Winkler, "A perceptual distortion metric for digital color video," *Proc. SPIE*, vol. 3644, no. 1, pp. 175–184, 1999.

[11] A. Watson, J. Hu, and J. McGowan, III, "Digital video quality metric based on human vision," *J. Electron. Imaging*, vol. 10, no. 1, pp. 20–29, 2001.

[12] J. Lubin and D. Fibush, "Sarnoff JND vision model," Tech. Rep. T1A1, vol. 5, 1997, pp. 97–612.

[13] B. Olshausen and D. Field, "How close are we to understanding V1?" *Neural Comput.*, vol. 17, no. 8, pp. 1665–1699, 2005.

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

[15] D. Tolhurst and J. Movshon, "Spatial and temporal contrast sensitivity of striate cortical neurones," *Nature*, vol. 257, pp. 674–675, Oct. 1975.

[16] S. Friend and C. Baker, "Spatio-temporal frequency separability in area 18 neurons of the cat," *Vis. Res.*, vol. 33, no. 13, pp. 1765–1771, Sep. 1993.

[17] M. Morrone, M. Di Stefano, and D. Burr, "Spatial and temporal properties of neurons of the lateral suprasylvian cortex of the cat," *J. Neurophysiol.*, vol. 56, no. 4, pp. 969–986, 1986.

[18] Y. Nie and K. Ma, "Adaptive rood pattern search for fast block-matching motion estimation," *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1442–1449, Dec. 2002.

[19] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Proc. ICIP*, Oct. 2006, pp. 2945–2948.

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

[21] A. K. Moorthy and A. C. Bovik, "Perceptually significant spatial pooling techniques for image quality assessment," in *Proc. SPIE*, Jan. 2009, pp. 1–11.

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

[23] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. Asilomar Signals Syst. Comput.*, Nov. 2003, pp. 1398–1402.

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

[25] S. Zhu and K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Trans. Image Process.*, vol. 9, no. 2, pp. 287–290, Feb. 2000.

[26] D. Rouse and S. Hemami, "Understanding and simplifying the structural similarity metric," in *Proc. ICIP*, 2008, pp. 1188–1191.
[27] *Final Report From the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment* [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI
[28] 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.
[29] K. Seshadrinathan and A. C. Bovik, "Unifying analysis of full reference image quality assessment," in *Proc. ICIP*, 2008, pp. 1200–1203.
[30] 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.
[31] D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*. Boca Raton, FL: CRC Press I LLC, 2004.
[32] I. Richardson, *H. 264 and MPEG-4 Video Compression*. West Sussex, England: Wiley, 2003.
[33] *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification*, ITU-T Rec. H.264/ISO/IEC 14 496-10 AVC, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, 2003.
[34] *H.264/AVC Software Coordination*. (2007) [Online]. Available: http://iphome.hhi.de/suehring/tml/