Perceptual Evaluation of Psychovisual Rate-Distortion Enhancement in Video Coding

Zhengfang Duanmu, Kai Zeng, Zhou Wang and Mahzar Eisapour
Dept. of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada

Abstract

Psychovisual rate-distortion optimization (Psy-RD) has been used in the industrial video coding practice as a tool to improve perceptual video quality. It has earned significant popularity through the wide spread of the open source x264 video encoders, where the Psy-RD option is employed by default. Nevertheless, little work has been dedicated to validate the impact of Psy-RD optimization on perceptual quality, so as to provide meaningful guidance on the practical usage and future development of the idea. In this work, we build a database that contains Psy-RD encoded video sequences at different strength and bitrates. A subjective user study is then conducted to evaluate and compare the quality of the Psy-RD encoded videos. We observe that there is considerable agreement between subjects' opinions on the test video sequences. Unfortunately, the impact of Psy-RD optimization on video quality does not appear to be encouraging. Somewhat surprisingly, the perceptual quality gain of Psy-RD ON versus Psy-RD OFF cases is negative on average. Our results suggest that Psy-RD optimization should be used with caution. Further investigations show that most state-of-the-art full-reference objective quality models correlate well with the subjective experiment results overall. But in terms of the paired comparison between Psy-RD ON and OFF cases, the false alarm rates are moderately high.

Introduction

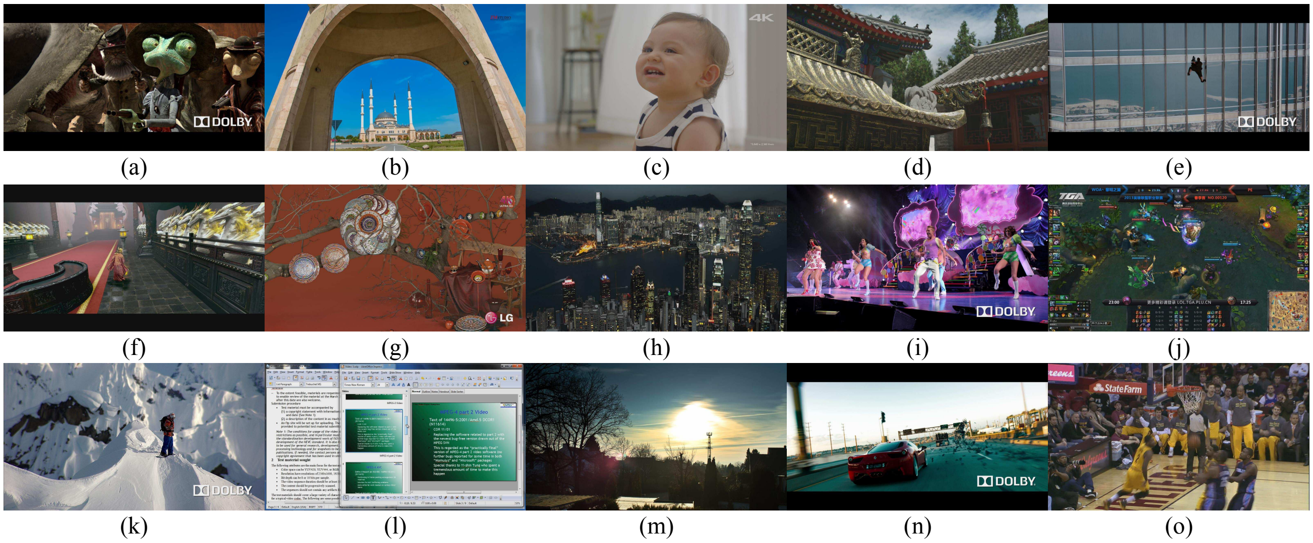
Video codecs are primarily characterized in terms of the throughput of the channel and the perceived distortion of the reconstructed video. A fundamental issue in video coding is to assign the available bis in an optimal way so as to obtain the best trade-off between the rate and perceived distortion. The process used to achieve this objective is commonly known as Rate Distortion Optimization (RDO). In practice, distortion models such as Sum of Absolute Difference (SAD) and Peak Signal-to-Noise Ratio (PSNR) are used in most actual implementations. However, these models do not correlate well with the perceptual video quality. Psychovisual rate-distortion optimization has been proposed to match perceived visual quality better by replacing the default distortion measure by more sophisticated objective models. Pyschovisual optimization has been a heavily studied research topic in academia [1, 2, 3, 4, 5, 6]. In the industry, DivX Labs made one of the first attempts to introduce psychovisual enhancement into their Dr. DivX [7] codec based on visual property of Just Noticeable Difference (JND). Dr. DivX analyzes each frame and concentrate on areas that are believed to be more noticeable to the human eyes. There are two optional settings available for Psychovisual Enhancements, namely shaping and masking. Shaping attempts to enhance fine details in the texture and mask differences between the source and encoded video in complex textures, making them less noticeable. Masking uses a slightly different algorithm, whereby each block in the frame and the surrounding blocks are analyzed such that the psychovisual enhancement introduces minimal artifacts. Another psychovisual optimized rate-distortion optimization, namely Psy-RD [8], was included in the x264 encoder and has been widely used in the video coding community. The philosophy behind Psy-RD is that the human eyes prefer the image to have similar complexity rather to look similar to the original image. In other words, humans would rather see a somewhat distorted but detail-rich block than a non-distorted but blurry block. Therefore, a tradeoff between the high frequency component of images and extra artifacts in the low frequency component region was introduced in the x264 encoder to increase the complexity of image especially when it is heavily compressed. This is very different from the traditional image quality assessment philosophy which considers the human visual contrast sensitivity variations across frequency and tend to lean towards sacrificing the quality in the high energy components. In the past few years, many users of x264 encoder have claimed that there was perceptual quality improvement when the Psy-RD optimization was turned on. However, to the best of our knowledge, the performance of Psy-RD has not been systematically studied. In [9], 5 video sequences with different x264 encoder settings including Psy-RD are tested, and the conclusion is that Psy-RD achieves a marginal gain to the default setting. So far, no extensive test that consists of different bitrates and Psy-RD strength was conducted, and more importantly, systematic subjective verification is completely missing. Consequently, whether the Psy-RD option should be turned on and what strength should be used to achieve the best visual quality is still unknown.

The purpose of this work is firstly to build a database that contains Psy-RD encoded videos at different Psy-RD strength and bitrate levels. Subjective experiment is then conducted using the test sequences and the mean opinion score (MOS) of each sequence is obtained. The results can be used to 1) study the human behaviors in evaluating the Psy-RD encoded video and analyze the impact of different Psy-RD settings; 2) test the performance of existing objective video quality assessment algorithms in predicting the subjective quality under psychovisual rate-distortion enhancement and explore potential ways to improve them.

Video Database and Subjective Quality Assessment

Video Database

Fifteen original high-quality videos of 1280×720 resolution are selected to cover diverse content types including humans, plants, natural sceneries, man-made architectures and computer-



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

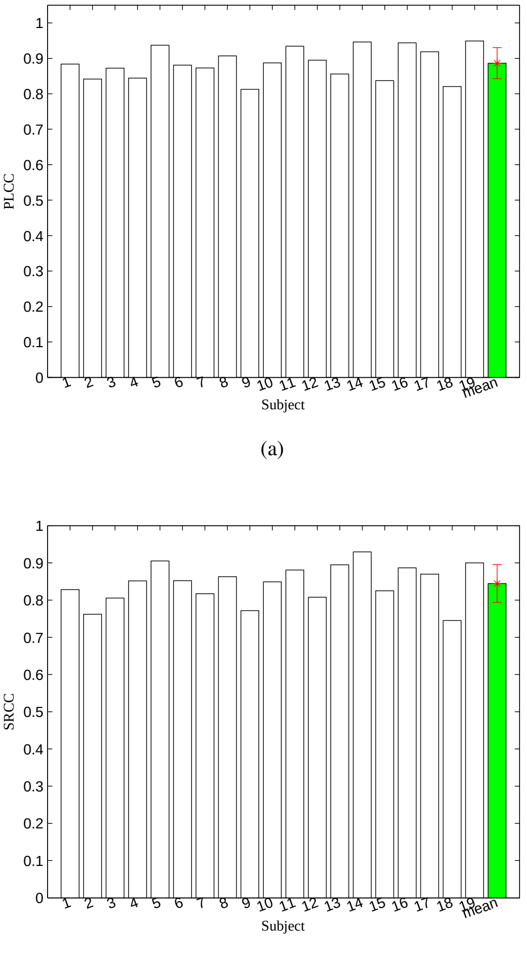
Figure 1. *Screen shot from each of the input source video clips used in the subjective study. (a) Animation. (b) Argun. (c) Baby. (d) China. (e) Climbing. (f) DaNaoTianGong. (g) Food. (h) HongKong. (i) KatyPerry. (j) LoL. (k) Skii. (l) SlideEditing. (m) TimeElapse. (n) Transformer. (o) ZapHighlights.*

synthesized sceneries. Fig. 1 shows the screen shots for all test videos. All videos have a duration of 10 seconds and a frame rate of 25 frames per second (fps). We created 16 test sequences from each of the reference sequences using x264 encoder at four different bit rates (250 kbps, 500 kbps, 950 kbps and 1300 kbps) and at four Psy-RD strength (0, 0.6, 1.0 and 2.0) to cover the commonly used working range of the Psy-RD tool.

Subjective Experiment

The subjective experiment was conducted on a PC with Intel(R) Core(TM) i7-2600 dual 3.40GHz CPU. All videos are displayed at their actual pixel resolution on an LCD monitor at a resolution of 2560 × 1600 pixel with Truecolor (32bit) at 60Hz. The monitor was calibrated in accordance with the recommendations of ITU-T BT.500 [10]. The test environment was setup as a normal indoor office workspace with ordinary illumination level. A customized subjective video quality assessment experiment program was used to render the videos on the screen and collect subjective opinion scores. During the test, the order of the video clips was randomized and thus different for each subject.

A total of 20 naive observers, including 12 males and 8 females aged between 20 and 40, participated in the subjective experiment. For each video clip, the subject was asked to give an integer score that best reflects the perceptual quality. For each subject, the whole study takes about one hour, which is divided into two sessions with a 7-minute breaks in-between to minimize the influence of fatigue effect. The score ranges from 0 to 100, where 0 denotes the worst quality and 100 the best. The choice of a 100-point continuous scale as opposed to a discrete 5-point ITU-R Absolute Category Scale (ACR) has advantages: expanded range, finer distinctions between ratings, and demonstrated prior efficacy [11].



(a)

(b)

Figure 2. *PLCC and SRCC between individual subject ratings and MOS. Rightmost column: performance of an average subject.*

Analysis and discussion

Analysis of Subjective Data

After the subjective test, one outlier subject was removed based on the outlier removal scheme in [10], resulting 19 valid subjects. The final quality score for each individual video clip is computed as the average of subjective scores, namely the mean opinion score (MOS). Considering the MOS as the "ground truth", the performance of individual subject can be evaluated by calculating the correlation coefficient between individual subject ratings and MOS values for each video clips. Pearson linear correlation coefficient (PLCC) and Spearman's rank-order correlation coefficient (SRCC) are employed as the evaluation criteria [12]. Both criteria range between 0 to 1, where higher values indicate better performance. The performance of each subject is depicted in Fig. 2. The average performance across all individual subjects is also given in the rightmost columns of Fig. 2. It can be observed that in general the subjects agree with each other to a significant extent.

TABLE 1: MOS gain for different Psy-RD strength at different bitrate levels.

Target bitrate	Psy-RD strength		
Kbps	0.6	1	2
250	-0.3867	-3.1867	-3.2367
500	-1.6233	-0.9500	-4.6433
950	-0.2200	-1.7267	-4.0867
1300	-0.6033	-0.9767	-0.8833

TABLE 2: Actual mean bitrate comparison

Target bitrate	Psy-RD strength			
Kbps	0	0.6	1	2
250	253.44	258.06	260.29	264.84
500	493.00	499.40	502.95	510.02
950	953.85	962.03	965.81	969.68
1300	1319.50	1306.73	1314.50	1317.09

To evaluate the effectiveness of Psy-RD on different video content, we plot the likelihood of Psy-RD on improving quality for each video content in Fig. 3 where likelihood is computed as the percentage of Psy-RD improving quality minus 0.5. It can be seen that although Psy-RD degrades the quality for most of the video content, especially for the videos with low spatial and temporal complexities, *e.g.*, Baby, DaNaoTianGong, and Skii, it tends to improve the quality of certain videos that contain complex spatial and temporal activities, *e.g.*, Animation, China, LoL, and Transformer. To the best of our knowledge, this phenomenon has not been explicitly reported in the literature. The reason behind is not fully understood but is worth deep investigation.

From the subjective test results, we have the following observations: 1) Table 1 lists the average MOS gain achieved by turning Psy-RD at different strength and different bitrate levels using Psy-RD OFF as the anchor. Consistent MOS loss is observed from the table, which means turning Psy-RD on would on average hurt the overall perceptual quality of videos. The larger the Psy-RD strength, the stronger the negative impact. 2) Psy-RD tends to increase the actual bitrate of videos as shown in Table 2. The larger of the Psy-RD strength, the larger the bitrate of the encoded video. 3) The impact of Psy-RD is content dependent. We observe that Psy-RD often improves the quality of complex-scene videos, and the gain peaks at Psy-RD strength 0.6. On the other hand, the quality of most of videos is hurt by Psy-RD, especially for the videos with low spatial and temporal complexity. Overall, as a psychovisual enhancement tool, Psy-RD should be used with caution because turning Psy-RD on not only increases bitrates, but may also introduce annoying artifacts that have significant negative impact on the perceptual quality.

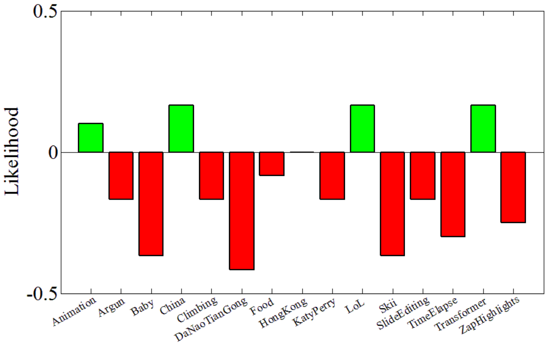


Figure 3. *Likelihood of Psy-RD on improving quality for different video content. Likelihood computed as percentage minus 0.5.*

Performance of Objective VQA Models

We tested 9 full-reference and 1 no-reference objective VQA models, including PSNR, VSNR [13], WSNR [14], S-SIM [15], MSSSIM [16], SSIMplus [17], VIF [18], STMAD [19], VQM [20], and BRISQUE [21]. Four criteria are employed for performance evaluation by comparing MOS and objective VQA models. Some of the criteria are included in previous tests carried out by the video quality experts group [12]. Other criteria are adopted in previous study [22]. These evaluation criteria are: 1) PLCC after a nonlinear modified logistic mapping between the subjective and objective scores [22]; 2) SRCC; 3) Mean absolute error (MAE) after the non-linear mapping; and 4) Root mean square after the non-linear mapping. Among the above metrics, PLCC, MAE and RMS are adopted to evaluate prediction accuracy, and SRCC is employed to assess prediction monotonicity [12]. A better objective VQA measure should have higher PLCC and SRCC while lower RMS and MAE values. Table 3 summarizes the evaluation results. It can be observed that state-of-the-art no-reference approach does not provide adequate predictions of the Psy-RD optimized videos. Several full-reference IQA models (S-SIM, MSSSIM, SSIMplus, VIF, and VQM) performs reasonably and almost equally well, although their computational cost is drastically different, ranked from SSIMplus, SSIM, MSSSIM, VIF to VQM, from the lowest to the highest. The scenario can also be observed from the scatter plots of the VQA algorithms in Fig. 4. Nevertheless, the good overall correlations between the subjective scores and model predictions do not necessarily mean that the objective models can well predict the exact impact of Psy-RD optimization on individual videos, for which deeper investigations are desirable.

To determine whether an objective VQA model can be used to automate the decision process that whether Psy-RD should be turned on in video coding, we further performed a false alarm test. Specifically, we compute the probability that each objective VQA model is not consistent with MOS in the direction of quality variation caused by Psy-RD optimization. Table 4 summarizes the evaluation results, which are somewhat disappointing because state-of-the-art VQA models do not seem to provide adequate predictions on the directions of quality variations caused by Psy-RD. Even the model with the best performance has an average false alarm rate higher than 0.3, which suggests a more accurate VQA model should be developed to evaluate the performance of Psy-RD optimization.

TABLE 3: Performance of objective VQA models

VQA model	SRCC	Computation Time (normalized based on PSNR)
PSNR	0.8468	**1**
VSNR [13]	0.2245	13.80
WSNR [14]	0.8502	16.39
SSIM [15]	0.8975	3.64
MSSSIM [16]	0.8859	18.36
SSIMplus [17]	**0.9168**	**1.78**
VIF [18]	0.9066	438.32
STMAD [19]	0.8133	673.67
VQM [20]	0.9079	43.26
BRISQUE [21]	0.3199	38.42

TABLE 4: False alarm rates of objective metrics on the performance improvement/degradation

VQA model	Psy-RD strength		
	0.6	1.0	2.0
PSNR	0.4833	0.4000	0.3333
VSNR[13]	0.4833	0.3833	0.4667
WSNR[14]	0.4833	0.4000	0.3333
SSIM[15]	0.4666	0.4167	0.3167
MSSSIM[16]	0.4666	0.4000	0.3167
SSIMplus[17]	0.4666	0.4000	0.3333
VIF[18]	0.4167	0.3834	0.2833
STMAD[19]	0.4333	0.3667	0.3333
VQM[20]	0.4667	0.4167	0.3333
BRISQUE[21]	0.5000	0.5833	0.6334

Conclusion and Future Work

We make one of the first attempts dedicated to investigating the perceptual effect of Psy-RD optimization in video coding. A database of Psy-RD optimized videos was created, followed by subjective experiment and data analysis. Our results are somewhat surprising, suggesting that Psy-RD optimization on average not only increases the consumption of bitrate and the computational resources, but also degrades the perceptual quality of encoded videos. Several objective VQA measures provide reasonable overall quality predictions, but may not precisely predict the effect of Psy-RD option on the perceived quality of individual videos. Our current study, though overall negative, does not necessarily lead to the conclusion that Psy-RD types of optimization are meaningless for perceptual video coding, but rather suggests that deeper investigations on both objective video quality assessment and perceptually inspired video coding are desirable to achieve consistent perceptual coding gain in real-world applications.

References

[1] Mai, Z.-Y., Yang, C.-L., and Xie, S.-L., "Improved best prediction mode (s) selection methods based on structural similarity in h. 264 i-frame encoder," in [*Proc. IEEE Int. Conf. System, Man and Cybernetics*], **3**, 2673–2678 (Oct. 2005).

[2] Yang, C.-L., Leung, R.-K., Po, L.-M., and Mai, Z.-Y., "An SSIM-optimal h. 264/avc inter frame encoder," in [*Proc. IEEE Int. Conf. Intelligent Computing and Intelligent Systems*], **4**, 291–295, IEEE (2009).

[3] Huang, Y.-H., Ou, T.-S., Su, P.-Y., and Chen, H. H., "Perceptual rate-distortion optimization using structural similarity index as quality metric," *IEEE Trans. Circuits and Systems for Video Tech.* **20**(11), 1614–1624 (2010).

[4] Chen, H. H., Huang, Y.-H., Su, P.-Y., and Ou, T.-S., "Improving video coding quality by perceptual rate-distortion optimization," in [*Proc. IEEE Int. Conf. Multimedia and Expo*], 1287–1292 (July 2010).

[5] Wang, S., Rehman, A., Wang, Z., Ma, S., and Gao, W., "SSIM-motivated rate-distortion optimization for video coding," *IEEE Trans. Circuits and Systems for Video Tech.* **22**(4), 516–529 (2012).

[6] Wang, S., Rehman, A., Wang, Z., Ma, S., and Gao, W., "Perceptual video coding based on SSIM-inspired divisive normalization," *IEEE Trans. Image Processing* **22**(4), 1418–1429 (2013).

[7] DivX LLC, "Psychovisual Enhancements," (Feb. 2007).

[8] Shikari, D., "Psychovisually optimized rate-distortion optimization," (Aug. 2008).

[9] Graphics, M. and Lab, M., "Mpeg-4 avc/h.264 video codecs comparison 2010 - Appendixes, url = http://www.compression.ru/video/codec-comparison/h264-2010/appendixes.html, year = "2010",."

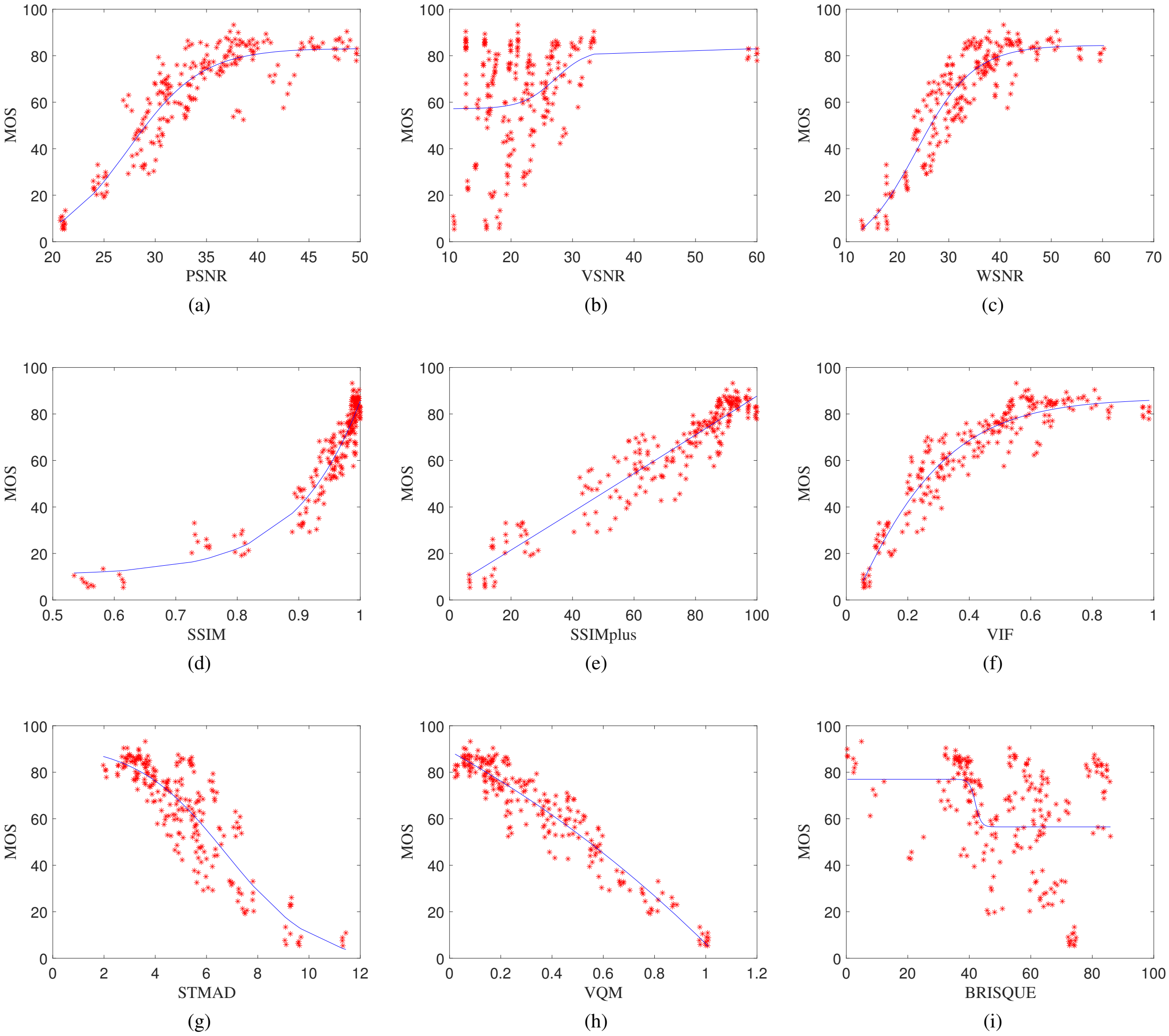
[10] ITU-R BT.500-12, "Recommendation: Methodology for the subjective assessment of the quality of television pictures," (Nov. 1993).

[11] Seshadrinathan, K., Soundararajan, R., Bovik, A. C., and Cormack, L. K., "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Processing* **19**(6), 1427–1441 (2010).

[12] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," (Apr. 2000).

[13] Chandler, D. M. and Hemami, S. S., "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Processing* **16**(9), 2284–2298 (2007).

[14] Mitsa, T. and Varkur, K. L., "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," in [*Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*], **5**, 301–304 (Apr. 1993).



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 4. *MOS versus predicted video quality by different objective models.*

[15] Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing* **13**(4), 600–612 (2004).

[16] Wang, Z., Simoncelli, E. P., and Bovik, A. C., "Multiscale structural similarity for image quality assessment," in [*Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*], **2**, 1398–1402 (Nov. 2003).

[17] Rehman, A., Zeng, K., and Wang, Z., "Display device-adapted video Quality-of-Experience assessment," in [*IS&T/SPIE Electronic Imaging: Human Vision and Electronic Imaging*], **9394** (Feb. 2015).

[18] Sheikh, H. R. and Bovik, A. C., "Image information and visual quality," *IEEE Trans. Image Processing* **15**(2), 430–444 (2006).

[19] Vu, P. V., Vu, C. T., and Chandler, D. M., "A spatiotemporal most-apparent-distortion model for video quality assessment," in [*Proc. IEEE Int. Conf. Image Proc.*], 2505–2508 (Sept. 2011).

[20] Pinson, M. H. and Wolf, S., "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcasting* **50**(3), 312–322 (2004).

[21] Mittal, A., Moorthy, A. K., and Bovik, A. C., "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Processing* **21**(12), 4695–4708 (2012).

[22] Sheikh, H. R., Sabir, M. F., and Bovik, A. C., "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing* **15**(11), 3440–3451 (2006).

Author Biography

Zhengfang Duanmu received the B.A.Sc degree in Electrical and Computer Engineering from the University of Waterloo (2015), where he is currently pursuing the M.A.Sc degree in electrical and computer engineering. His research interests lie in perceptual image processing and quality of experience.

Kai Zeng received the Ph.D. degree in Electrical and Computer Engineering from University of Waterloo (2013), where he is currently a Post-Doctoral Fellow. His research interests include computational video and image communication and processing. Dr. Zeng was a recipient of IEEE Signal Processing Society student travel grant at the 2010 and 2012, and prestigious 2013 Chinese Government Award for Outstanding Students Abroad.

Zhou Wang is currently a Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image processing, coding, and quality assessment; computational vision and pattern analysis; multimedia communications; and biomedical signal processing. He has more than 100 publications in these fields with over 30,000 citations (Google Scholar). He is a Fellow of IEEE and Canadian Academy of Engineering, and a recipient of an NSERC Steacie Fellowship, two IEEE Signal Processing Society Best Paper Awards, and a Primetime Engineering Emmy Award.

Mahzar Eisapour received the B.Math degree in Computer Science from the University of Waterloo (2009). Her research interest lies in image processing.