PERCEPTUALLY-FRIENDLY RATE DISTORTION OPTIMIZATION IN HIGH EFFICIENCY VIDEO CODING

Sima Valizadeh, Panos Nasiopoulos and Rabab Ward

Department of Electrical and Computer Engineering, University of British Columbia
Vancouver, Canada

ABSTRACT

We propose the employment of a perceptual video quality metric in measuring the distortion in the High Efficiency Video Coding (HEVC) Standard. The mean square error presently used as quality metric is not a good measure to use, as it poorly correlates with human perception. Integration of a video quality metric based on the characteristics of the Human Visual System (HVS) inside the rate distortion optimization procedure is expected to improve the compression efficiency of the video coding. In this paper, the PSNR-HVS measure is used in the rate distortion optimization process. The compression efficiency of the proposed approach is compared to that used by HEVC, the recent video coding standard. Simulations prove that the proposed approach yields higher compression efficiency and provides better visual quality.

Index Terms— Perceptual video coding, rate distortion optimization (RDO), human visual system (HVS), PSNR-HVS, high efficiency video coding (HEVC)

1. INTRODUCTION

Recently, the ITU-T Visual Coding Experts Group (VCEG) and the ISO-IEC Moving Picture Experts Group (MPEG) have developed an improved video coding standard called High Efficiency Video Coding (HEVC) [1]. HEVC achieves improved compression efficiency by introducing new tools such as variable size for coding units and prediction units, extended intra prediction modes, advanced motion vector prediction, block merge method, and adaptive loop filtering. The HEVC test model provides 35.4% bit rate savings in comparison to H.264/AVC with the same image fidelity [2].

To improve the efficiency of video coding while keeping the quality high, the utilization of the characteristics of the Human Visual System (HVS) can play an important role. As humans are the ultimate consumers and judges of the video content, the subjective evaluation of the video quality is what matters the most. Utilization of video quality metrics that are based on the Human Visual System and their integration within the video encoder will improve the quality of the compressed video.

For the H.264 standard, many methods have been recently proposed to integrate the properties of the human visual system into some aspects of this standard. The studies in [3]-[4] attempt to introduce the properties of the HVS into the quantization process of video coding in H.264. As the sensitivity of the HVS differs for different frequencies, a frequency weighting scheme has been used in the quantization process of H.264.

In the H.264 standard also, the SSIM metric has been used in inter frame prediction and mode selection [5]-[9]. Also, SSIM has been integrated into the rate control of H.264 for intra frame coding [10]. In [11]-[14], a perceptual mode selection scheme based on SSIM is integrated into the rate distortion optimization of H.264.

The High Efficiency Video Coding, HEVC, standard has extended the H.264 intra prediction modes, making its mode selection process different from H.264. Also, the variable size Coding Units and variable size Prediction Units in HEVC, were not available in H.264. In this study, our proposed method modifies the prediction and mode selection in HEVC.

In a more recent study, SSIM has been used in the rate distortion optimization of the HEVC standard [15]. Our study shares the same goal of increasing the coding efficiency of the video coding. We integrate a perceptual video quality metric that is based on HVS, named PSNR-HVS [16], inside the video encoder. PSNR-HVS, compared to PSNR, shows higher correlation with subjective video quality evaluations. PSNR-HVS has the advantage of being easily adoptable to blocks of data and is not too computationally complex to be integrated inside the video encoder.

The rest of the paper is organized as follows: Section 2 describes rate distortion optimization in the high efficiency video coding. Section 3 explains the integration of HVS-based video quality metrics in the rate distortion optimization. Section 4 compares our experimental results with the reference software. Finally, conclusions are made in section 5.

2. HEVC RATE DISTORTION OPTIMIZATION

The job of the encoder is to select the coding parameters in a way that results in the best coding efficiency. In video coding, the bitrate is minimized for a certain fixed distortion, or the distortion is minimized for a certain fixed rate:

This work was supported by the NPRP grant # NPRP 4-463-2-172 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

$$\min(D) \quad subject\ to\ R \le R_T \tag{1}$$

Minimization is done over the coding parameters, D is distortion and R is the number of bits required to signal coding parameters. This minimization process is formulated via a non-negative Lagrange multiplier λ in the Rate Distortion Optimization (RDO) [17] process:

$$\min \quad J = (D + \lambda \times R) \tag{2}$$

In HEVC, the rate distortion minimization is done in four stages [18]. In the first stage, the mode is decided for each coding unit. Afterwards, intra prediction mode estimation is performed. In the third stage, motion estimation is carried out and finally the last stage is quantization.

For coding unit mode decision, distortion is measured by Sum of Square Error (SSE) and λ_{mode} is defined as [19]:

$$\lambda_{mode} = \alpha \times w_k \times 2^{((QP-12)/3.0} \tag{3}$$

Alpha = 1.0_Clip3(0.0, 0.5, 0.05 * number _ of _ B _ frames) for referenced pictures

Alpha = 1.0 for non-referenced pictures

$$Clip3(x,y,z) = \begin{cases} x; z < x \\ y; z > y \\ z; otherwise \end{cases} \tag{4}$$

Measuring the distortion by SSE and minimizing it in the rate distortion optimization process, leads to better PSNR in the coded video. However, the PSNR has been shown to have limited correlation with subjective tests. Various other video quality metrics have been developed to better represent how subjects evaluate the video quality.

3. INTEGRATION OF HVS-BASED VIDEO QUALITY METRICS IN THE RATE DISTORTION OPTIMIZATION

In psychophysics, the human visual system is modeled by a transfer function. This transfer function is used to identify a system representing the visual cortex. Electrophysiological experiments showed that the visual cortex cells are sensitive to spatial frequency bands and orientation [20]-[21]. Contrast sensitivity and masking are the two main concepts that govern the visual perception. Contrast sensitivity accounts for the perception of single wavelength, or stimuli. Masking quantizes the interactions between several stimuli. Human eye is less sensitive to higher spatial frequency than to lower.

Perceptual video quality metrics are classified into methods in the pixel domain as well as the frequency domain [22]. Discrete cosine transform (DCT), wavelet transform and Gabor filter banks are used in the frequency domain methods. Pixel domain methods use local gradient changes around a pixel or extract visual features based on computational models of the low level vision. One of the earliest perceptual video quality metrics in the frequency domain consists of filtering and masking processes [23]. Motion Picture Quality Metric (MPQM) [24] is based on a multi-channel model of human spatio-temporal vision. Digital Video Quality (DVQ) [25] estimates the local contrast based on the ratio of DCT amplitude to DC component. From the local contrast, Just Noticeable Differences (JNDs) are estimated. An extension of DVQ, uses the fact that the human eyes' sensitivity to spatio-temporal patterns decrease with high spatial and temporal frequencies. This metric uses a spatial contrast sensitivity (SCS) matrix for static frames and SCS raised to a power for dynamic frames [26]. Another perceptual video quality metric in the frequency domain is based on the Wavelet Transform [27]. Motion-based Video Integrity Evaluation (MOVIE) models the response characteristics of the middle temporal visual area with separable Gabor filter banks [28].

PSNR-HVS [16] is a full reference video quality metric which takes into account the characteristics of the human visual system. One of the characteristics of the HVS is that its sensitivity decreases at high spatial frequencies. PSNR-HVS is defined as:

$$PSNR - HVS = 10 \log\left(\frac{255^2}{MSE_{HVS}}\right) \tag{5}$$

$$MSE_{HVS} = K \sum_{i=1}^{I-7} \sum_{j=1}^{J-7} \sum_{m=1}^{8} \sum_{n=1}^{8} ((X[m,n]_{ij} - X[m,n]_{ij}^{e}) T_C[m,n])^2$$

where K=1/[(I-7)(J-7)64]. I, J are the image width and height. X_{ij} is the DCT coefficient of an 8×8 image block with its upper left corner at (i, j). X_{ij}^{e} is the DCT coefficient of the corresponding block in the original image. T_c is a matrix adopted from the JPEG quantization table proposed in the JPEG [29].

PSNR-HVS considers a window size of 8×8. It considers the DCT of the window; A matrix of correcting factors adopted from the JPEG quantization table gives more weights to lower frequency coefficients. It has been shown that PSNR-HVS has higher correlation with the subjective results than PSNR does [16]. PSNR-HVS has desirable properties that allow its application to coding units in the video compression based on HEVC.

Moreover, PSNR-HVS is a Distance Measure (DM). It is important to note that not all metrics satisfy the requirements of distance measures. PSNR is a distance measure in the pixel domain and it satisfies the following properties:

- $D[A, A'] = f(|a_{11} - a'_{11}|, |a_{12} - a'_{12}|, \dots, |a_{nn} - a'_{nn}|)$
-
$\forall i, j, x_{11}, \dots, x_{nn}: \ |a_{ij}| < |a'_{ij}| \Rightarrow$
$f(|x_{11}|, \dots, |a_{ij}|, \dots, |x_{nn}|) \le f(|x_{11}|, \dots, |a'_{ij}|, \dots, |x_{nn}|)$

where A and A' are two matrices; a_{ij} and a'_{ij} are the elements at (i, j) index; D measures the distance between matrices A and A'. The distance measure, D, is defined by function f. Triangular inequality does not imply that the second condition is met.

PSNR-HVS satisfies these two conditions in the frequency domain. However, not all metrics belong to this class.

Only video quality metrics that satisfy requirements of being a Distance Measure can be used to measure distortion in the rate distortion optimization process.
In the rate distortion optimization process, λ acts as a knob that controls the trade-off between rate decreases versus distortion increases. Distortion measurements with a video quality metric changes this trade-off based on the range of output values. New optimal value of λ needs to be determined to have the best quality of video while minimizing the required bitrate. In this study, a scaling factor is used to denote the relationship between the new λ and the λ used in HEVC.

4. EXPERIMENTAL RESULTS

To validate the efficiency of the proposed approach, PSNR-HVS is integrated into the HEVC reference software HM9.2. For the test video sequences, we used MPEG standard videos as summarized in Table 1. All test video sequences are in the YCbCr 4:2:0 format. The length of each sequence is 10 seconds.

Sequences	Resolution	Frame Rate	Total Frames
BQ Square	416× 240	60	600
Race Horse	416 × 240	30	300
Party Scene	832 × 480	50	500
BQ Terrace	1920 × 1080	60	600

Table 1. Test video sequences

Quantization Parameters (QP) of values 22, 27, 32 and 37 have been tested. These QP values was selected by MPEG during HEVC standardization process.
Various scaling factors are tested to find the optimal trade-off between the rate and distortion. In our tests, the Lagrangian multiplier is modified by a scaling factor as:

$$\lambda_{proposed} = \mathrm{c} \times \lambda_{mode} \qquad (6)$$

where $\lambda_{proposed}$ is the proposed Lagrangian multiplier in our approach, λ_{mode} is the Lagrangian multiplier in HEVC as defined in equation (3) and c is a scaling factor. Scaling factors of 1.6, 4, 5.6, 8, 12 and 16 have been tested in this paper. Scaling factors below 1.6 leads to high bitrates. Scaling factors above 16 results in low quality in the compressed video. Fig. 1 shows the Rate Distortion performance for video sequence BQSquare. The quality of the video is measured with Mean PSNR-HVS. PSNR-HVS has more correlation with subjective tests than PSNR [16]. Fig. 1 shows that our proposed method achieves higher quality at the same bitrate compared to HEVC reference software over the full range of Quantization Parameter (QP) values. For BQSquare video sequence, scaling factor of 5.6 leads to higher bitrate saving compared to scaling factor of 1.6 or 16. As Fig. 1(b) shows the improvements become more significant at higher bitrates.
The average bitrate difference between the proposed and reference rate-distortion curves is referred to Bjontegaard's Delta (BD) Rate [30]. A cubic polynomial approximation is derived using four data points (quality and bitrate points). The difference between the two curves is integrated in the horizontal direction for BD Rate. BD Rate measures the average bitrate savings by the proposed approach compared to the HEVC reference software.
The bitrate savings depend on the scaling factor. There is a trade-off between minimizing distortion and the required bitrate. This trade-off is controlled by the scaling factor applied to λ. As the scaling factor increases, more emphasis is given to the bitrate in the minimization process; while lower scaling factors emphasizes more on lower distortion values. Bitrate savings of the proposed approach for video BQSquare based on different scaling factors is summarized in Table 2.

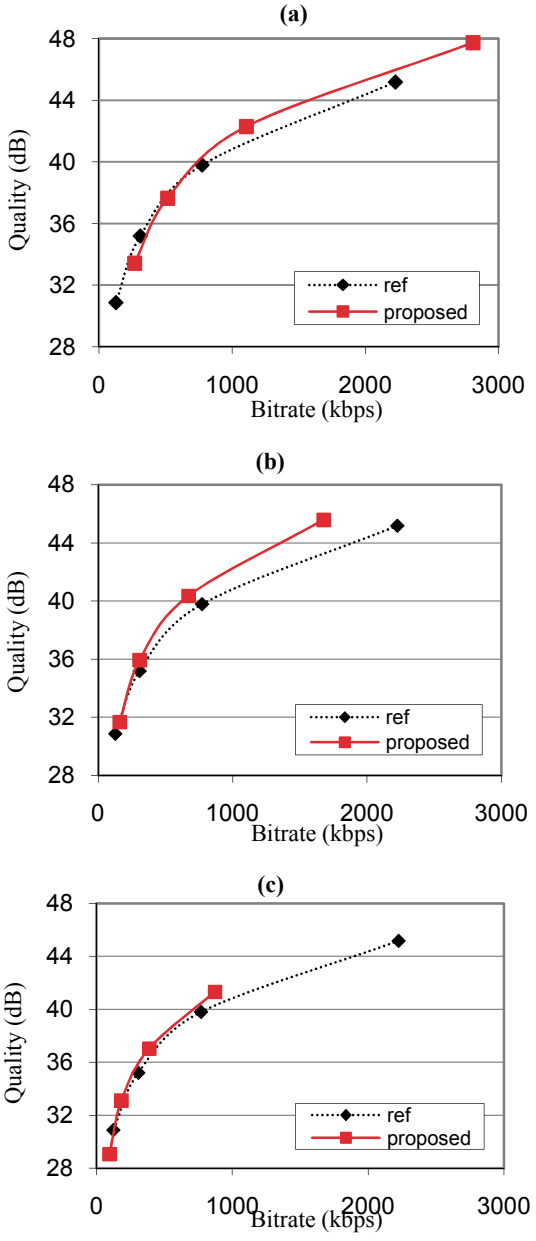

Fig. 1 Rate-Distortion Curves for BQSquare video sequence for three different scaling factors. Quality is measured with Mean PSNR-HVS. (a) Scaling factor=1.6. (b) Scaling factor=5.6. (c) Scaling factor=16.

Scaling factor	1.6	5.6	16
BD Rate	-5.8%	-21.3%	-15.9%

Table 2. Bitrate saving of the proposed approach compared to the reference HEVC software for different scaling factors for the BQSquare video sequence.

To find the optimal scaling factor, bitrate savings versus scaling factors are plotted in Fig. 2 for different video sequences. For the video sequence RaceHorses, scaling factor 5.6 show the most bitrate saving. However, scaling factor 8 shows highest amount of bitrate saving for video sequences of BQSquare, BQTerrace and PartyScene. Thus, scaling factor 8 is selected for our proposed approach.

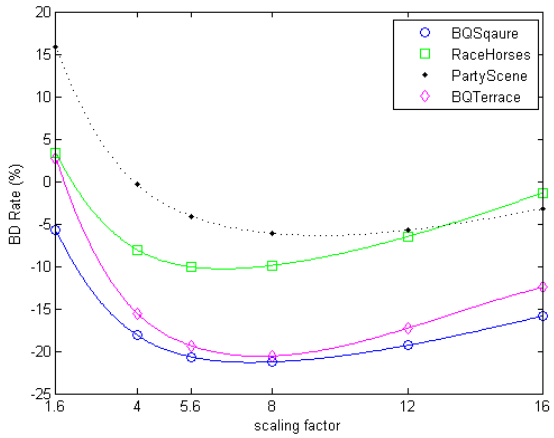

Fig. 2 Bitrate savings versus the scaling factor for different video sequences

Based on Fig. 2, for the video BQSquare, the amount of bitrate saving changes very minimally as the scaling factor sweeps the range of 4 to 12. The sensitivity of the bitrate savings to the scaling factor can be analyzed based on Fig. 2. The slop of line tangent to the curves increases monotonically by moving towards lower or higher scaling factors.

Table 3 shows the bitrate and their corresponding video quality for the test video sequences with HEVC and the proposed approach. Quantization parameters of 22, 27, 32 and 37 are used in this table. These results show that the proposed approach can deliver the same video quality at a lower bitrate. Thus, the proposed approach improves the coding efficiency of the video coding.

	Reference HEVC		**Proposed**	
QP	Bitrate (kbps)	Quality (dB)	Bitrate (kbps)	Quality (dB)
22	2225.06	45.18	1362.82	44.20
27	772.30	39.79	541.38	39.34
32	308.49	35.19	240.54	35.10
37	126.62	30.87	120.57	31.09

Table 3. Bitrate and quality of the proposed method along with the HEVC reference software for the test video sequence BQSquare.

Integration of the quality metric inside the encoder achieves higher compression efficiency at the expense of more complexity. In this study, the complexity that is introduced is taking DCT of blocks and applying a matrix of frequency weighting. The encoding time of the proposed approach along with the reference HEVC software are reported in Table 4. It is worth noting that the same QP will result in different bitrates with the reference software and the proposed approach as reported in Table3. For instance, encoding the BQSquare video with the reference software results in 2,225 kbps for QP=22. The same QP value of 22, results in the bitrate of 1,362 kbps when encoded with the proposed approach. Thus, different bitrates also contribute to different encoding times although the QP is the same.

	Encoding Time (s)	
QP	Reference HEVC	Proposed
22	3083.78	3218.59
27	2352.79	2533.15
32	1853.21	2020.79
37	1583.61	1764.24

Table 4. Comparison between encoding time of the proposed algorithm and HEVC reference software

The results in Table 5 show the bitrate savings for each of the video sequences. BQSquare results in the most BD Rate followed by BQ Terrace. These video sequences have less motion compared to the other two test videos.

Sequences	$\Delta rate$
BQ Square	-21.3%
Race Horses	-5.2%
Party Scene	-9.3%
BQ Terrace	-20.6%
Average	-14.1%

Table 5. Rate reduction of the proposed method compared to HEVC reference software (over QPs of 22, 27, 32 and 37)

Based on Table 5, the proposed approach results in the average bitrate saving of 14.1% over the test sequences.

5. CONCLUSION

In this paper, we proposed to integrate a video quality metric inside HEVC video coding standard. PSNR-HVS is a quality metric based on the human visual system. We used PSNR-HVS as the measure of distortion in the rate-distortion optimization in the encoding process. The proposed approach was tested on different standard video sequences. The results show that our proposed scheme requires on average 14.1% less bitrate with the same perceived video quality compared to HEVC. In our future work, we will investigate the dependency of the optimal Lagrangian multiplier on different values of the quantization parameter (QP) in our proposed approach.

REFERENCES

[1] G. J. Sullivan, J. Ohm, Woo-Jin Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 22, pp. 1649-1668, 2012.

[2] J. Ohm, G. J. Sullivan, H. Schwarz, Thiow Keng Tan and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards—Including High Efficiency Video Coding (HEVC)," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 22, pp. 1669-1684, 2012.

[3] Z. Chen and C. Guillemot, "Perceptually-Friendly H.264/AVC Video Coding Based on Foveated Just-Noticeable-Distortion Model," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 20, pp. 806-819, 2010.

[4] J. Chen, J. Zheng and Y. He, "Macroblock-Level Adaptive Frequency Weighting for Perceptual Video Coding," *Consumer Electronics, IEEE Transactions on,* vol. 53, pp. 775-781, 2007.

[5] Z. Mai, C. Yang, K. Kuang and L. Po, "A novel motion estimation method based on structural similarity for H.264 inter prediction," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing,* vol. 2, Feb. 2006, pp. 913-916, 2006.

[6] C. Yang, R.Leung, L. Po and Z. Mai, "An SSIM-optimal H.264/AVC inter frame encoder," in *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on,* 2009, pp. 291-295.

[7] C. Yang, H. Wang and L. Po, "Improved inter prediction based on structural similarity in H.264," in *Signal Processing and Communications, 2007. ICSPC 2007. IEEE International Conference on,* 2007, pp. 340-343.

[8] Z. Mai, C. Yang, L. Po, and S. Xie, "A new rate-distortion optimization using structural information in H.264 I-frame encoder," *in Proc. ACIVS* 2005, pp. 435–441.

[9] Z. Mai, C. Yang and S. Xie, "Improved best prediction mode(s) selection methods based on structural similarity in H.264 I-frame encoder," in Proc. IEEE Int. Conf. *Sys. Man Cybern.,* May 2005, pp. 2673-2678 Vol. 3.

[10] B. H. K. Aswathappa and K. R. Rao, "Rate-distortion optimization using structural information in H.264 strictly intra-frame encoder," in *System Theory (SSST), 2010 42nd Southeastern Symposium on,* 2010, pp. 367-370.

[11] S. Wang, A. Rehman, Z. Wang, S. Ma and W. Gao, "Perceptual Video Coding Based on SSIM-Inspired Divisive Normalization," *Image Processing, IEEE Transactions on,* vol. 22, pp. 1418-1429, 2013.

[12] S. Wang, A. Rehman, Z. Wang, S. Ma and W. Gao, "SSIM-Motivated Rate-Distortion Optimization for Video Coding," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 22, pp. 516-529, 2012.

[13] T. Ou, Y. Huang and H. H. Chen, "SSIM-Based Perceptual Rate Control for Video Coding," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 21, pp. 682-691, 2011.

[14] C. Yeo, H. Li Tan and Y. H. Tan, "On Rate Distortion Optimization Using SSIM," *Circuits and Systems for Video Technology, IEEE Transactions on,* vol. 23, pp. 1170-1181, 2013.

[15] A. Rehman and Z. Wang, "SSIM-inspired perceptual video coding for HEVC," in *Multimedia and Expo (ICME), 2012 IEEE International Conference on,* 2012, pp. 497-502.

[16] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, M. Carli, New full-reference quality metrics based on HVS, *Proc. of the Second International Workshop on Video Processing and Quality Metrics*, Scottsdale, USA, 2006, 4 p.

[17] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *Signal Processing Magazine, IEEE,* vol. 15, pp. 74-90, 1998.

[18] V. Sze, M. Budagavi, G. Sullivan, *High Efficiency Video Coding (HEVC)*, Springer, 2014.

[19] K. McCann, B. Bross, W. Han, I. Kim, K. Sugimoto, G. Sullivan, High Efficiency Video Coding (HEVC) Test Model 13 (HM 13) Encoder Description, Joint Collaborative Team on Video Coding (JCT-VC), Document JCTVC-O1002, Geneva, Oct. 2013.

[20] R. L. DeValois and K. K. DeValois, *Spatial Vision.* Oxford University Press, 1988.

[21] J. G. Daugman, "Spatial visual channels in the Fourier plane," Vision Res., vol. 24, pp. 891-910, 1984.

[22] S. Chikkerur, V. Sundaram, M. Reisslein and L. J. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison," *Broadcasting, IEEE Transactions on,* vol. 57, pp. 165-182, 2011.

[23] F. Lukas and Z. L. Budrikis, "Picture Quality Prediction Based on a Visual Model," *Communications, IEEE Transactions on,* vol. 30, pp. 1679-1692, 1982.

[24] Van den Branden Lambrecht, Christian J and O. Verscheure, "Perceptual quality measure using a spatiotemporal model of the human visual system," in *Electronic Imaging: Science & Technology,* 1996, pp. 450-461.

[25] A. B. Watson, J. Hu and J. F. McGowan, "Digital video quality metric based on human vision," *Journal of Electronic Imaging,* vol. 10, pp. 20-29, 2001.

[26] F. Xiao, "DCT-based video quality evaluation," *Final Project for EE392J,* vol. 769, 2000.

[27] C. Lee and O. Kwon, "Objective measurements of video quality using the wavelet transform," *Optical Engineering,* vol. 42, pp. 265-272, 2003.

[28] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *Image Processing, IEEE Transactions on,* vol. 19, pp. 335-350, 2010.

[29] G. Wallace, "The JPEG still picture compression standard," *Comm. of the ACM*, vol. 34, No.4, 1991.

[30] G. Bjontegaard, "Calculation of average PSNR difference between RD curves," *in Proc. 13th Meeting ITU-T Q.6/SG16 VCEG* , Austin, TX, Apr. 2001.