# Improved Adaptive Video Delivery System Using a Perceptual Pre-processing Filter

Louis J. Kerofsky, Rahul Vanam, and Yuriy A. Reznik
InterDigital Communications, Inc., 9710 Scranton Road, San Diego, CA 92121 USA
E-mail: {louis.kerofsky, rahul.vanam, yuriy.reznik}@interdigital.com

***Abstract*—In this paper, we present an adaptive video delivery system that uses a perceptual pre-processing filter tuned to the viewing conditions of the receiver. The filter receives a number of parameters of the reproduction setup that includes viewing distance, display pixel density, ambient contrast, display peak luminance, etc. A model of visual contrast sensitivity is used in the filter design to remove spatial oscillations that are invisible under the specified viewing conditions. Removing such oscillations simplifies the video content leading to bitrate savings in the video encoding without causing any visible alterations of the content. Experiments demonstrate the bitrate savings our filter can yield compared to conventional encoding methods that are not tailored to specific viewing conditions. Subjective viewing tests confirm the bitrate savings come without reduction in visual quality.**

***Index Terms*—Contrast sensitivity function, Barten CSF model, perceptual coding, adaptive coding**

## I. Introduction

Design of an efficient video delivery system is challenging. Compression is the primary tool used in a video delivery system to address the vast bandwidth needed to transmit raw video. Effective operation requires reducing the bandwidth needs while minimizing the impact on perceptual quality experienced by a viewer. The visual sensitivity of a viewer is influenced by several factors of the display and viewing conditions such as distance between viewer and screen, pixel density of the screen, ambient illuminance, etc.

Parameters of the reproduction setup are not known in conventional video coding and delivery systems, and they are often assumed to be within a certain range (e.g., viewing distance equal to 3–4×height of the display) or worst case values are assumed. However, as illustrated in Figure 1, it is conceivable to design an adaptive system that would measure such viewing parameters dynamically and pass them back to the transmitter. In turn, the transmitter may use this information for effective encoding of video for a particular reproduction setting. For example, as shown in Figure 1, such customization of encoding can be accomplished by using a perceptual pre-processing filter.

We propose a pre-processing filter suitable for use in such a system. The following visual effects are exploited that are known to affect the visibility of a signal:

- *contrast sensitivity function* (CSF) [1] - relationship between frequency and contrast sensitivity thresholds of human vision.
- *eccentricity* - rapid decay of contrast sensitivity as angular distance from gaze point increases.


Fig. 1. Architecture of adaptive video delivery system. The pre-processing filter is used to remove spatial oscillations invisible under current reproduction setup.

- *Oblique effect* - higher visibility of horizontal and vertical lines compared to diagonal ones.

The above phenomena are well known, and have been used in the field of image processing. For example, CSF models have been used in quality assessment methods such as Visible Differences Predictor (VDP) [2], SQRI metric [1], S-CIELAB [3], etc. The oblique effect has been incorporated in some of these CSF models [1], [2]. Previously suggested applications of eccentricity included coding with eye-tracking feedback, foveal coding [4], etc.

However, our application is different. We do not use eye tracking, and our filter only receives global characteristics of the viewing setup, such as viewing distance, contrast, etc. Also, our goal is not to identify or measure visual differences, but to remove spatial oscillations that are invisible under given viewing conditions. Removing such oscillations simplifies video content, thereby leading to more efficient encoding without causing visible alterations of the content.

In our prior work [5], [6], we developed a perceptual pre-processing filter that considers viewing parameters such as viewing distance, display pixel density, display size, and ambient contrast ratio. In this paper, we consider additional viewing parameters that affect visual perception that include peak luminance of the display, surround reflectance, and device reflectance. These parameters are used to adapt a CSF model, which provides a more accurate response of the human visual system to different viewing parameters.

Through experiments, we demonstrate that the use of our filter can yield significant bit rate savings compared to conventional encoding methods that are not tailored to specific viewing conditions. We also compare our filter to a conventional low-pass filter with cutoff frequency set to match the visual acuity limit under the same viewing conditions. We show that our filter outperforms such conventional design by an appreciable margin. Subjective viewing experiments were conducted to confirm the visual model based design. The results of subjective testing confirm the bitrate savings come without statistically significant visual image differences.

This paper is organized as follows. In Section II we explain

details of our filter design. In Section III we study performance of this filter. In Section IV we offer conclusions.

## II. Design of a perceptual pre-filter

### A. Background

The concept of Contrast Sensitivity Function (CSF) has been studied by a variety of authors to explain response of the human visual system. We use the CSF in the design of our pre-filter. The CSF describes the spatial frequency perception of the human visual system (HVS), and provides a relationship between contrast sensitivity and spatial frequency in cycles-per-degree (cpd) [7]. Contrast sensitivity is defined as inverse of contrast threshold. The Michelson's contrast is defined as

$$CT = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = \frac{amplitude(I)}{mean(I)}, \tag{1}$$

where $I_{max}$ and $I_{min}$ are maximum and minimum intensities of oscillation. We use the mathematical model of the CSF developed by Barten [8], which incorporates a range of parameters. In particular, we use the following parameters of the viewing conditions:

a) Viewing distance. The viewing distance is used in computing the spatial frequency in cpd, thereby incorporating viewing distance into the CSF. The spatial frequency $f$ (in cpd) of a sinusoidal grating with cycle length $n$ pixels can be computed as: $f = 0.5\left[\arctan\left(\frac{n}{2\,d\,\rho}\right)\right]^{-1}$, where $\rho$ is the display pixel density (in ppi), and $d$ is the distance between viewer and the screen (in inches).

b) Surround luminance. It is known that the contrast sensitivity depends on the contents of the visual field surrounding the region of a stimulus and not merely the area of gaze. We use a model in [8] that describes the impact of the surround effect on the CSF. This model scales the CSF depending upon the ratio between surround luminance and the luminance of the stimulus, and the object size.

c) Ambient contrast. Reflection of ambient light from the display surface can considerably lower the effective contrast ratio. This equivalently places a floor on the contrast sensitivity of images projected on the screen. The contrast sensitivity in the presence of ambient light can be computed as

$$\begin{aligned} s_A &= \frac{I_{max} + I_{min} + 2I_R}{I_{max} - I_{min}} \\ &= \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} + \frac{2I_R}{I_{max} - I_{min}} = \frac{1}{CT_0} + s_{min}, \end{aligned} \tag{2}$$

where $CT_0$ is the contrast threshold in dark defined in Equation (1), and $I_R$ is the intensity of light reflected from the display. $I_{max}$ and $I_{min}$ were defined earlier. The offset $s_{min}$ can be computed directly from the ambient contrast ratio $CR_A$ as

$$s_{min} = \frac{1}{CR_A} = \frac{\frac{L_{peak}}{CR_0} + \frac{r_D}{\pi} I_A}{L_{peak} + \frac{r_D}{\pi} I_A}, \tag{3}$$

where $CR_0$ is the device native contrast ratio, $r_D$ is the device reflectivity, $I_A$ is the ambient irradiance, and $L_{peak}$ is the peak display brightness.

TABLE I
Filter input parameters.

| Parameter | Tablet use case |
|---|---|
| Viewing distance ($d$) | 12", 15", 20", 25", 30" |
| Ambient illuminance ($I_A$) | 50 , 500 , 10000 lux |
| Resolution ($X_0$) | 1080P ($1920 \times 1080$) |
| Peak luminance ($L_{Peak}$) | 200 $cd/m^2$ |
| Device native CR ($CR_0$) | 900:1 |
| Surround reflectance ($r_S$) | 20% |
| Device reflectivity ($r_D$) | 7.7% |
| Device density ($\rho$) | 264ppi |

### B. Perceptual Pre-filter Design

The perceptual pre-filter operates using a spatially adaptive filter which removes image oscillations which are not visible under the current viewing conditions. The pre-filter operates independently on each video frame as an image. A block diagram of our filter is shown in Figure 2. The viewing parameters that are provided as input to the pre-filter are listed in Table I. The pre-filter is described next using the "Kodak03" input image, shown in Fig 2(a), as an example.

*1) Linear space conversion and black level adjustment:* The input video/image is first converted to linear color space followed by extraction of a luminance channel $y$. To model display response, we further raise the black level based on ambient contrast:

$$y' = \alpha + (1 - \alpha)y, \tag{4}$$

where $\alpha = 1/CR_A$, and $CR_A$ is defined in Eq (3). Figure 2 (b) shows the result of this operation.

*2) DC estimation:* We incorporate the eccentricity effect in our local DC calculation. We estimate local DC values by applying a Gaussian low pass filter to the luminance image. We select filter parameter $\sigma$ to achieve a cutoff of about $\frac{1}{4}$ cpd. This achieves smooth averaging within a region that can be captured by foveal vision. Figure 2 (c) illustrates the local DC estimate after low pass filtering. We denote the DC value at location $(i, j)$ as $\mathrm{DC}_{ij}$.

*3) Estimation of contrast sensitivity:* The difference image is obtained by taking the absolute difference between the estimated DC and luminance images. The envelope of amplitude fluctuations is obtained by further applying a max filter. The length of max filter is selected to be identical to the support length of our final adaptive low-pass filter. Figure 2 (d) illustrates the amplitude envelope image. Let $\mathrm{amplitude}_{ij}$ be the amplitude at location $(i, j)$. The contrast sensitivity at location $(i, j)$ is subsequently computed as

$$x_{ij} = \frac{\mathrm{DC}_{ij}}{\mathrm{amplitude}_{ij}}. \tag{5}$$

*4) Estimation of object luminance:* The Global DC of a given frame is computed from the black level adjusted image. In most natural videos the DC varies slowly across time. Therefore, we apply a temporal low pass filter across the global DC values of current frame and past frame. The object luminance ($L_o$) is then estimated by taking a product of the peak display luminance ($L_{peak}$) with the filtered global DC.

*5) Highest visible frequency estimation:* Using the obtained contrast sensitivity values $x_{ij}$ we next estimate the highest

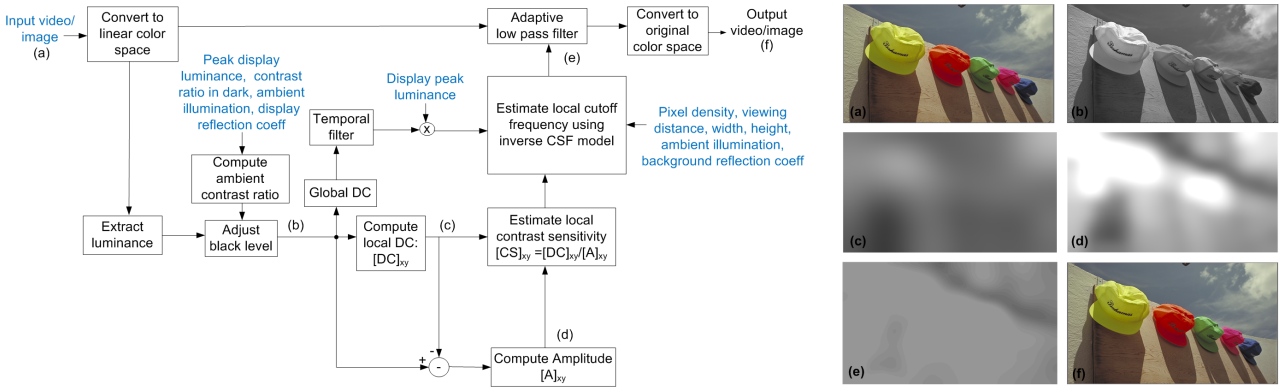


Fig. 2. Block diagram of Perceptual Prefilter architecture. The input parameters to the filter are in blue font. (a) "Kodak03" test image, (b) black level adjusted luminance, (c) local DC estimate, (d) amplitude envelope estimate, (e) cutoff frequency, and (f) filtered output image.

spatial frequencies which will be visible. An approximate inverse of the Barten CSF model [8] is derived considering the upper branch of the CSF as shown in Figure 3(a). The $LambertW$ function [9] is used in our inverse calculation. This inverse function is used to map the sensitivity lower bound to an upper bound on the highest visible frequency, and is given as:

$$f_c(x_{ij}) = \sqrt{\frac{\text{LambertW}(\frac{2DA^2e^{2DB}}{(C+1)x_{ij}^2})}{2 \cdot D}} - B, \tag{6}$$

where $f_c$ is the spatial cutoff frequency in cpd, and $A, B, C, D$, and $E$ are defined as follows:

$$A = \frac{5200E}{\sqrt{0.64}}, B = \frac{1}{0.64}\left(1 + \frac{144}{X_0^2}\right),$$
$$C = \frac{63}{L_o^{0.83}}, D = 0.0016\left(1 + \frac{100}{L_o}\right)^{0.08},$$
$$E = \exp\left(-\frac{\ln^2(\frac{L_s}{L_o} \cdot (1 + \frac{144}{X_0^2})^{0.25}) - \ln^2((1 + \frac{144}{X_0^2})^{0.25})}{2 \cdot \ln^2(32)}\right), \tag{7}$$

where $L_o$ is the object luminance, $X_0$ is the object size in visual degrees, and $L_s$ is the surround luminance. The factor $E$ models the impact of surround effect on the CSF [8]. The surround luminance $L_s$ can be estimated from ambient irradiance ($I_A$) and surround reflectance ($r_S$) as $L_s = \frac{I_A r_S}{\pi}$, where $r_S$ is typically chosen to be 20% [10].

The maximum cutoff frequency is $f_{max} = f_c(s_{min})$, since contrast sensitivity is lower bounded by $s_{min}$. Figure 3(b) illustrates $f_{max}$ vs. ambient illumination for sample conditions of Table I, where $s_{min}$ and associated $f_{max}$ were calculated for each ambient illumination level using Equations (3) and (6), respectively.

*6) Oblique filter operation:* We incorporate the oblique effect in our pre-filter by using the approach in [11]. This yields a cutoff frequency of $f_c$ along the cardinal directions and lower cutoff frequencies along diagonal directions, with a minimum cutoff frequency of $0.78 f_c$ along $45^o$.

## III. Experimental Setup and Results

### A. Compression performance

Four 1080p, 25fps test videos are used in Table II. The x264 encoder [12] is used to produce High-Profile H.264/AVC-compliant bitstreams. To produce encodings of both original and filtered videos with closest possible amounts of distortions, we use constant QP rate control with the same QPs applied in encoding both original and pre-filtered videos. For each test video, we determine QPs at which an unfiltered video would yield bitrates of 15, 10, and 5 Mbps, respectively, and they are listed in Table II. For comparison with our adaptive filter, we use a conventional low-pass filter, with cutoff frequency selected as $f_c^{uniform} = f_{max}$.

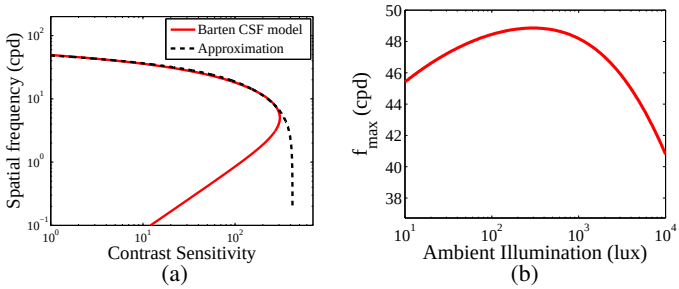


Fig. 3. (a) Computing cutoff frequency using an approximation of the inverted CSF model [8]. (b) Highest visible frequency $f_{max}$ vs. ambient illumination for sample conditions of Table I.

The test videos are filtered considering typical viewing conditions for a tablet, listed in Table I. They include 5 viewing distances ranging from 12" to 30", and three ambient illuminances ranging from 50 lux (dimly lit room) to 10000 lux (bright daylight). Both the perceptual and uniform pre-filtered videos are encoded using QPs chosen from Table II. Resulting bitrates are used to make two comparisons: (1) bitrate of the proposed pre-filter vs. unfiltered encodings, and (2) bitrate of the uniform pre-filter vs. unfiltered encodings.

The resulting plots averaged across test videos at different viewing distances, ambient illumination, and bitrates are shown in Figure 4. Across different test conditions, our filter yields higher bitrate savings over the uniform filter; the maximum bitrate savings ranges between $[30.4\%, 9.6\%]$ for our filter, and $[22.9\%, 4.2\%]$ for the uniform filter.

As expected, with longer viewing distance both filters yield higher bitrate savings. Bitrate savings for both filters decrease with the following order of ambient illumination: 10000 lux, 50 lux, and 500 lux. This can be explained using Figure 3 (b), which illustrates that ambient illumination of 500 lux yields

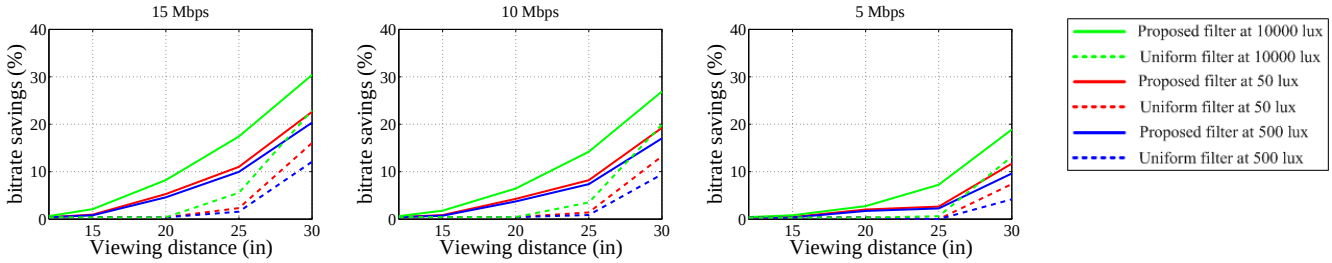


Fig. 4. Bitrate savings results averaged across test videos, for different bitrates, viewing distances and ambient illumination.

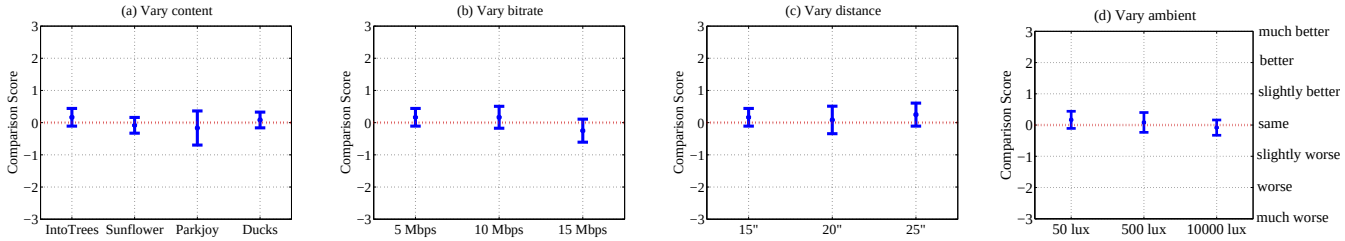


Fig. 5. Comparison Scores for varying test conditions comparing the proposed pre-filtering with the unfiltered sequence. (a) Content is varied, (b) bitrate is varied, (c) viewing distance is varied, and (d) ambient illumination is varied. The IntoTrees sequence is used for (b), (c), and (d). Mean and 95% confidence intervals of scores are shown.

TABLE II
SEQUENCES AND QPS USED FOR BIT-RATE COMPARISONS. ALL SEQUENCES ARE 1920 × 1080 AND 25 FPS.

| Sequence | QP | | |
|---|---|---|---|
| | 15 Mbps | 10 Mbps | 5 Mbps |
| IntoTrees [13] | 26 | 27 | 30 |
| Sunflower [14] | 18 | 19 | 22 |
| Parkjoy [13] | 34 | 36 | 40 |
| DucksTakeOff [13] | 35 | 38 | 42 |

maximum highest visible frequency, which results in a higher pass-band frequency for both filters, thereby yielding lower bitrate savings. The bitrate savings for 50 and 10000 lux can be explained similarly.

### B. Perceptual Quality

*1) Test Methodology:* For subjective testing we used a Microsoft Surface device with 10.6" screen, 1080p resolution, $L_{peak} = 200cd/m^2$, $CR_0$ = 1000:1, and $r_D$ = 5.8%. The Comparison Scale Method [15] was used to compare the perceptual quality of videos prepared with and without the use of our perceptual pre-filter. A Double Stimulus presentation was used to compare the original and pre-filter performance. Each test consisted of the following presentation: sequence A for 10s, grey for 3s, stimulus B for 10s, and then one repetition of all three videos. The subjects selected a vote from a 7-point scale described in [16], and the scale is illustrated along the vertical axes of Figure 5(d). Prior to the test session, a training session was conducted where the test methodology was described using a training stimulus. A total of 12 subjects were used ranging in age from 20's to 50's. Six among them were imaging experts, while the remaining were novices.

*2) Test Material:* Several comparisons were done varying single parameters: sequence content, target bitrate, viewing distance, and ambient illumination. Specific test criteria are listed in Table III. For the sequence variation tests, the sequences in Table II were used with the tablet conditions of 20" viewing distance and 500 lux ambient. In all other cases

TABLE III
TEST CONDITIONS

| Test | Bit-rate (Mbps) | Distance (") | Ambient (lux) |
|---|---|---|---|
| Vary content | 10 | 20 | 500 |
| Vary bitrate | 5, 10, 15 | 20 | 500 |
| Vary distance | 10 | 16, 20, 24 | 500 |
| Vary ambient | 10 | 20 | 50 , 500, 5000 |

the sequence IntoTrees was used to explore variation with: target bitrate, viewing distance, and ambient illumination, since it showed the most coding gain for the pre-filter across a variety of conditions. For each viewing condition listed in Table III, suitable pre-filter parameters were chosen for filtering the test videos. In each case, both the unfiltered and filtered videos were encoded using QPs selected from Table II. These two bitstreams were then visually compared using the test methodology described above, under suitable viewing conditions, to evaluate any perceptual difference. Opinion scores for each of the families of test are shown in Fig. 5. Results in Fig. 5 indicate that the perceptual quality of pre-filtered videos is statistically indistinguishable from unfiltered videos, under given viewing conditions.

## IV. CONCLUSION

We present a perceptual pre-processing filter that removes spatial oscillations invisible to a user under specified viewing conditions. Our pre-filter uses a number of viewing parameters that include viewing distance, ambient illumination, surround luminance, and display characteristics such as display reflectivity, density, and resolution. The pre-filter uses these parameters together with the Barten CSF model to determine filter cutoff frequency that is associated with highest visible frequency. Subjective tests show that our pre-filtered videos are visually close to the unfiltered videos for different video content, bitrates, viewing distances, and ambient illumination. Our pre-filter shows improved bitrate savings over the uniform pre-filter, and yields up to 30% savings over unfiltered encoding.

## REFERENCES


[1] P. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*. SPIE Press, 1999.

[2] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*. SPIE, 1992, pp. 2–15.

[3] X. Zhang, B. A. Wandell *et al.*, "A spatial extension of CIELAB for digital color image reproduction," in *SID international symposium digest of technical papers*, vol. 27. SID, 1996, pp. 731–734.

[4] A. C. Bovik, *Handbook of image and video processing*. Academic Press, 2005.

[5] R. Vanam and Y. Reznik, "Improving the efficiency of video coding by using perceptual preprocessing filter," in *Proc. Data Compression Conference*, 2013.

[6] R. Vanam and Y. A. Reznik, "Perceptual pre-processing filter for user-adaptive coding and delivery of visual information," in *Proc. 30th Picture Coding Symposium*, 2013, pp. 426–429.

[7] H. R. Wu, A. R. Reibman, W. Lin, F. Pereira, and S. S. Hemami, "Perceptual visual signal compression and transmission," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2025–2043, 2013.

[8] P. G. Barten, "Formula for the contrast sensitivity of the human eye," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2003, pp. 231–238.

[9] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, "On the lambertw function," *Advances in Computational mathematics*, vol. 5, no. 1, pp. 329–359, 1996.

[10] T. Fujine, Y. Yoshida, and M. Sugino, "The relationship between preferred luminance and TV screen size," in *Proc. SPIE*, vol. 6808, 2008, p. 68080Z.

[11] Y. Reznik and R. Vanam, "Improving the coding and delivery of video by exploiting the oblique effect," in *Proc. First IEEE Global Conference on Signal and Information processing*, Dec. 2013, pp. 775–778.

[12] "x264 encoder," http://www.videolan.org/developers/ x264.html.

[13] "The SVT high definition multi format test set," ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/.

[14] "Hdgreetings," http://www.hdgreetings.com/other/ecards-video/video-1080p.aspx.

[15] H. Wu and K. Rao, *Digital Video Image Quality and Perceptual Coding*. CRC Press, 2005.

[16] ITU-R BT 500-13, "Methodology for the subjective assessment of the quality of television pictures," *International Telecommunication Union, Geneva, Switzerland*, pp. 53–56, 2002.