PERCEPTUAL PRE-PROCESSING FILTER FOR ADAPTIVE VIDEO ON DEMAND CONTENT DELIVERY

Rahul Vanam, Louis J. Kerofsky, and Yuriy A. Reznik

InterDigital Communications, Inc., 9710 Scranton Road, San Diego, CA 92121 USA

E-mail: {rahul.vanam, louis.kerofsky, yuriy.reznik}@interdigital.com

ABSTRACT

We describe the use of perceptual pre-processing to reduce the bitrate needed for delivery of VOD content. The proposed system exploits the viewing conditions of a specific individual to remove image oscillations which are not visible by the user under the specific viewing conditions. The pre-processing uses parameters such as: viewing distance, pixel density, ambient illumination etc. A model of human visual system contrast sensitivity is used to remove oscillations in the image data which cannot be seen and need not be encoded. Experiments demonstrate significant bitrate savings compare to conventional encoding methods which do not exploit individual specifics.

Index Terms— Perceptual video coding, contrast sensitivity function, human visual system, Video on Demand, visual acuity.

1. INTRODUCTION

In most conventional video coding and delivery systems, viewing conditions are not known precisely and are not fully exploited in the video coding and delivery. In a related previous work we described how accurate feedback of the viewing conditions to an encoder/delivery system could be exploited for improved coding performance by adapting to the dynamics of a mobile user [1–4]. We consider here a living room user and a Video On Demand (VOD) application. In such a VOD application content may be encoded in a number of representations. The specific representation sent to a particular user may be selected in real-time to adapt to the specifics of the user as shown in Figure 1. Unlike our previous work, where we assumed full real-time feedback from the sensors on a mobile device for selecting the bitstreams, in this work, we use partial knowledge of the viewing condition and display parameters without assuming full dynamic real-time feedback. The viewing condition and display information may be determined or estimated in a number of ways as indicated in Figure 1, and could be based either on population statistics [5], information configured during a set up process or determined in real-time by a set-top box (STB) for instance. Time of day and geographic location may be exploited to estimate ambient lighting conditions. Such viewing conditions are variable per installation but relatively

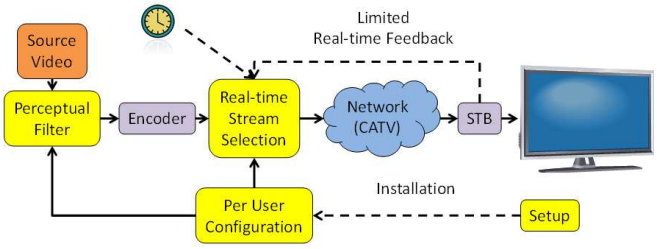

Fig. 1. Architecture of a system for adaptive Video on Demand (VOD) with proposed perceptual pre-processing filter.

constant at an individual site i.e. viewing distance, display size and resolution in a home are roughly constant in time. Adaptation to a single user's dynamic variation in time in the mobile use case is replaced by adaptation to individuals in an ensemble of users in the living room use case. Adaptation per individual in a VOD use case is most promising and the motivation of this work; however the underlying technique can be applied to general multicast applications as well where worst case conditions for an ensemble of users may be used. This paper discusses design of a pre-processing filter suitable for use in such a system. Our design exploits three basic phenomena of human vision [6–8]:

- *Contrast sensitivity function* (CSF) – relationship between frequency and contrast sensitivity thresholds of human vision.
- *Oblique effect* – phenomenon where human visual system is less sensitive to diagonally oriented spatial oscillations when compared to horizontal and vertical ones.
- *Eccentricity* – rapid decay of contrast sensitivity as angular distance from gaze point increases.

All three phenomena are well known, and have been used in image processing in the past. For example, CSF models have been used in quality assessment methods such as Visible Differences Predictor (VDP) [9], SQRI metric [8], S-CIELAB [10], etc. The oblique effect has been incorporated in some of these CSF models [9], [8]. Previously suggested applications of eccentricity included coding with eye-tracking feedback, foveal coding [6], etc.

Our application is different. We are not suggesting to use eye tracking, and our filter only uses 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. By removing such oscillations our filter simplifies video content, thereby leading to more efficient encoding without causing visible alterations of the content.

In this paper, we demonstrate that this perceptual filter yields significant bitrate savings compared to a conventional encoding scheme that is not tailored to specific viewing conditions. We also demonstrate a higher bitrate savings over our prior work [4] under similar viewing conditions.

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.

2. PERCEPTUAL PRE-FILTER DESIGN

2.1. Underlying principles

It is well known that the visual system has different sensitivity to different spatial frequencies. The design of our filter relies upon the Contrast Sensitivity Function (CSF) of human vision as described in [8] for example to exploit this effect. For our application, the visually relevant *cycles per degree* [cpd] is converted to cycles per pixel used in the filtering process. As exemplified in Figure 2(a), the spatial frequency f of a sinusoidal grating with cycle length of n pixels can be computed as:

$$f = \frac{1}{\beta}\,[\text{cpd}], \quad \beta = 2\arctan\left(\frac{n}{2\,d\,\rho}\right), \tag{1}$$

where ρ is the display pixel density (expressed in ppi), d is the distance between viewer and the screen (in inches), and β is the angular span of one cycle of the grating (in degrees). The variation in visual system sensitivity at different spatial frequencies is illustrated in Figure 2(b). The spatial frequency is expressed in cpd, and the contrast sensitivity is defined as the inverse of contrast thresholds. The Michelson's contrast of an oscillation is defined as:

$$C_T = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{\text{amplitude}(I)}{\text{mean}(I)} \tag{2}$$

where $I_{\max}$, $I_{\min}$ denote minimum and maximum intensities of an oscillation.

We must also note that the CSF characteristic is meaningful only for characterizing sensitivity to features localized in some small (about 2 degrees of viewing angle) spatial regions.

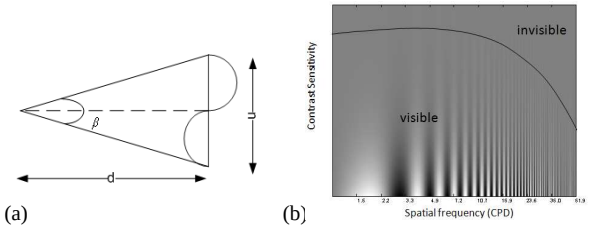

Fig. 2. (a) Illustration of the concept of spatial frequency. (b) Contrast sensitivity function (CSF) of human vision.

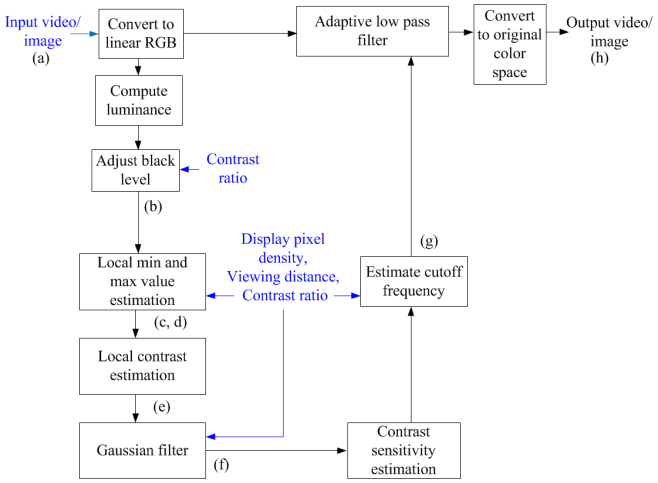

Fig. 3. Block diagram of our perceptual filter. The parenthesized letters refer to sub-figures in Figure 4. The inputs to the filter are in blue font.

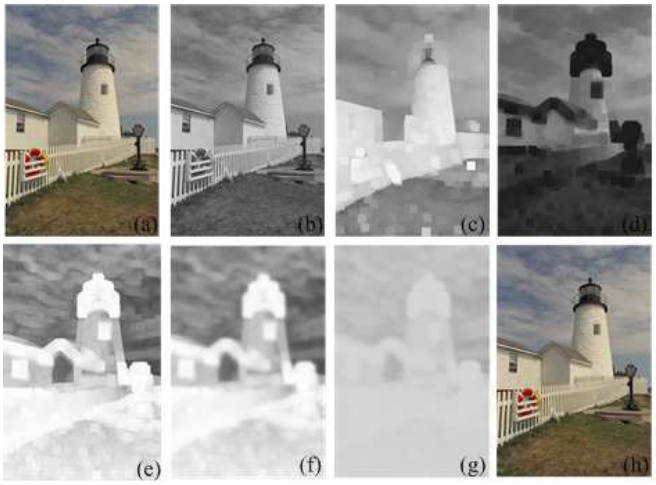

Fig. 4. (a) Kodak "k19" test image, (b) black level adjusted luminance, (c) max image, (d) min image, (e) contrast, (f) filtered contrast, (g) cutoff frequency map, and (h) filtered output image.

Due to the eccentricity of human vision, larger regions cannot be examined with the same acuity. This gives us an important cue on how to apply the CSF in our filter design.

2.2. Design of a perceptual pre-filter

A block diagram of our filter is shown in Figure 3. It is a spatial filter, processing each frame in the video sequence independently as an image. The inputs to the filter include input video/image, viewing distance between the display and the user, effective contrast ratio of the screen (for given ambient light and display brightness settings), and the display pixel density. We next explain the main processing steps in this design, and illustrate them using the Kodak "k19" input image, shown in Figure 4(a), as an example.

(a) Linear space conversion and black level adjustment:

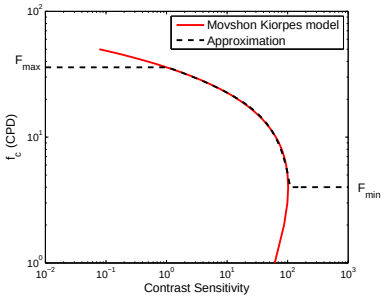

Fig. 5. Computing cutoff frequency using partially inverted CSF model [11].

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:

$$y' = \alpha + (1 - \alpha)y, \tag{3}$$

where $\alpha = 1/CR$, and CR is the effective contrast ratio of the display. Figure 4 (b) shows the result of this operation.

(b) Contrast estimation: In [4], local DC and amplitude estimates were used in estimating contrast. Instead we use local min and max values since they estimate local contrast with higher accuracy. We find the local min and max values of the black level adjusted image using a window of size 2 cpd; the min and max images are illustrated in Figures 4 (d) and (e), respectively. Using the min and max values the local contrast is estimated by computing the Michelson's contrast defined in Eq (1). The contrast image is filtered using a Gaussian low pass filter having 4 cpd length. This achieves smooth averaging within a region that can be captured by foveal vision. Figure 4 (f) illustrates the filtered contrast image.

(c) Cutoff frequency estimation: The contrast sensitivity at each point is computed by taking inverse of the filtered contrast value at the corresponding point. Let x_{ij} be the contrast sensitivity at location (i, j). Using the obtained contrast sensitivity values x_{ij} we next estimate the highest spatial frequencies which will be visible. For this, we employ the upper branch of the inverse CSF function, as shown in Figure 5. We further restrict results to range $[F_{\min}, F_{\max}]$, where F_{min} corresponds to a point where the CSF peaks, and F_{max} is the visual acuity limit.

For instance, when employing the Movshon and Kiorpes CSF model [11], this yields the following algorithm for computing the highest visible frequency $f_c(x_{ij})$:

$$\begin{aligned} f_c'(x_{ij}) &= -42.26 + 78.46{x_{ij}}^{-0.079} - 0.049{x_{ij}}^{1.08} \\ f_c(x_{ij}) &= \begin{cases} F_{min}, & f_c'(x_{ij}) < F_{min} \\ f_c'(x_{ij}), & F_{min} \leqslant f_c'(x_ij) \leqslant F_{max} \\ F_{max}, & f_c'(x_{ij}) > F_{max}. \end{cases} \end{aligned} \tag{4}$$

Figure 4 (e) shows cut-off frequencies computed using this formula. Darker colors imply heavier filtering.

(e) Filtering operation: Our adaptive filter incorporates the oblique effect phenomenon by strongly filtering frequencies oriented along diagonal direction (i.e., using $0.78 f_c(x_{ij})$ along 45^o) compared to those along horizontal and vertical directions (i.e., $f_c(x_{ij})$). We use the separable filter implementation described in [3] for performing this directional adaptive filtering. This filter operates in linear space, followed by conversion to the desired output color format. Figure 4 (h) illustrates the final filtered image.

Table 1. Sequences and QPs used at approximately 15 Mbps, 10 Mbps, and 5 Mbps. All videos are 1920×1080 and 25 fps.

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

3. EXPERIMENTAL SETUP AND RESULTS

In this section, we describe our experimental setup and present test results.

3.1. Video sequences and encoder settings

In our experiments we use four standard 1920x1080 videos listed in Table 1. We use the x264 encoder [12], configured to produce High-Profile H.264/AVC-compliant bitstreams. To produce encodings of both original and filtered content with closest possible amounts of distortions, we use constant QP rate control with the same QPs applied in encodings for both original and pre-filtered sequences. Specific choices of QP values that we selected for each sequence are shown in Table 1. These QPs were found to produce encodings of the original (non-filtered) sequences at approximately 15 Mbps, 10 Mbps and 5 Mbps rates, representing relevant operating points.

3.2. Viewing conditions

Conditions representing typical TV usage conditions [5] [15] are considered to evaluate the performance of our filter. We consider a 65" TV display in our test, and select the following 5 viewing distances expressed in heights of the display d = {3H, 4H, 4.5H, 5H, and 6H}. We select the following effective contrast ratios of the screen CR = {2:1, 5:1, 10:1, 100:1, 500:1}. The first corresponds to a situation in bright daylight, while the last assumes a dimly lit room at night.

3.3. Comparisons and verification

Given the above list of target viewing conditions we run our perceptual pre-filter and produce sequences pre-filtered for each combination of contrast and viewing distance parameters. We perform two types of comparison:

- Size of the encoded original vs. encoded sequence pre-filtered using [4], and
- Size of the encoded original vs. encoded sequence pre-filtered using our proposed method.

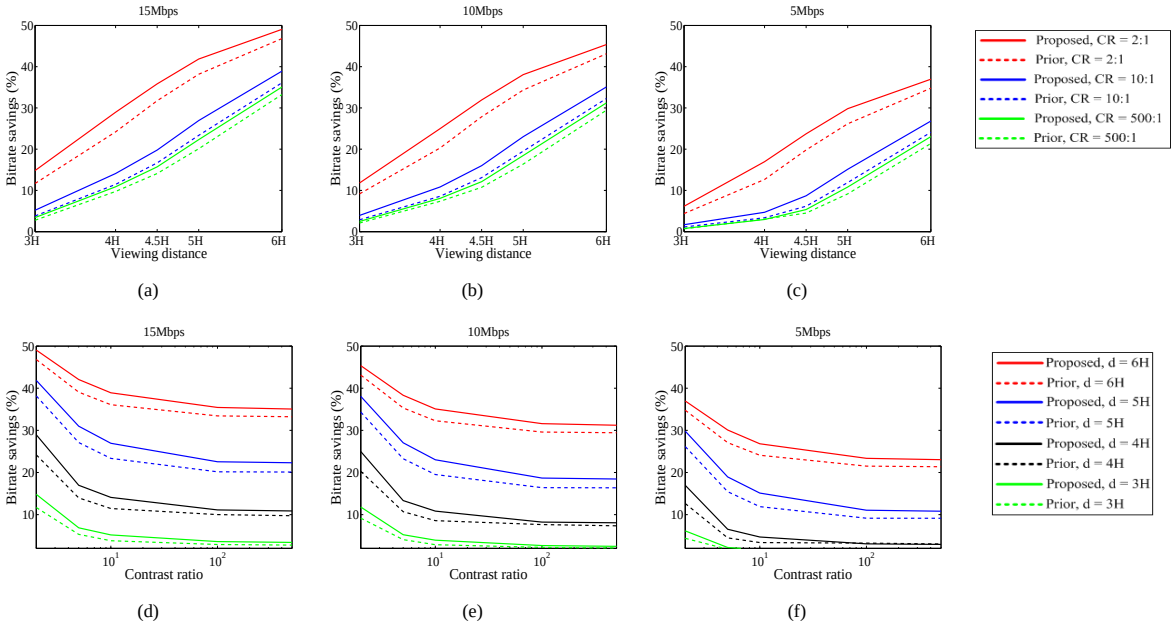

Fig. 6. Bitrate savings of proposed pre-filter and prior pre-filter [4] over unfiltered encoding averaged across test videos. Bitrate savings when varying viewing distance and setting display contrast ratio (CR) = 2:1, 10:1, and 500:1, at unfiltered encoding bitrates = 15 Mbps (a), 10 Mbps (b), and 5 Mbps (c). Bitrate savings when varying contrast ratio and setting viewing distance d = 3H, 4H, 5H, and 6H, for unfiltered encoding bitrates = 15 Mbps (d), 10 Mbps (e), and 5 Mbps (f).

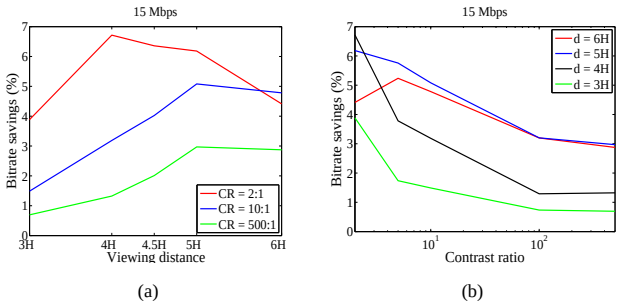

Fig. 7. Bitrate savings of proposed pre-filter over [4], averaged across test videos at 15 Mbps. (a) Results when CR = 2:1, 10:1, and 500:1; and (b) when d = 3H, 4H, 5H, and 6H.

As mentioned earlier, to ensure the same level of quality in encodings of original and filtered sequences we use the same encoder settings and same fixed QPs. In addition, we have also performed visual cross-checks of encodings with the goal of verifying that under specified viewing conditions both encoded original and encoded filtered sequences appear identical. Simultaneous double-stimuli viewing was performed by a panel of 5 viewers. We did this for 3 viewing distances (3H, 4H, and 5H) and effective contrasts of 100:1, and 10:1, and found no noticeable differences.

3.4. Results

We present average bitrate saving results achieved by the two pre-filters over original encoding in Figure 6. In Figures 6(a)-(c), we present bitrate savings vs. viewing distance when contrast ratio (CR) = $\{2{:}1,\ 10{:}1,\ 500{:}1\}$, and in Figures 6(d)-(f) we present bitrate savings vs. contrast ratio when viewing distance d = $\{3H,\ 4H,\ 5H,\ 6H\}$. As expected, both pre-filters yield higher bitrate savings with lower effective contrast ratios as shown in Figures 6(a)-(c). Also, longer viewing distance yields higher bitrate savings as shown in Figures 6(d)-(f). Relative bitrate savings are found to be larger at higher bitrates; we achieve maximum average bitrate savings of 37%, 45% and 50% at 5 Mbps, 10 Mbps, and 50 Mbps, respectively. We present average bitrate savings of our pre-filter over [4] at 15 Mbps only in Figure 7, since the trends are similar at 5 and 10 Mbps. From Figure 7(a), appreciable gains are seen at lower contrast ratios (CR $\leq$ 10). Figure 7(b) indicates that our pre-filter yields higher savings at d = 5H for most contrast ratios. Our pre-filter achieves a maximum bitrate savings of 6.8% over [4] at lowest CR and d = 4H.

4. CONCLUSION

We presented a perceptual pre-filter for adaptive VOD content delivery. This filter uses parameters of the reproduction setup, such as viewing distance, pixel density, and display contrast ratio, for removing spatial oscillations that are invisible under such viewing conditions. Through experiments, we demonstrate that our pre-filter yields up to 50% bitrate savings over a conventional encoding method that is not tailored to specific viewing conditions. Compared to [4], our pre-filter yields appreciable bitrate savings particularly in low-contrast regime.

5. REFERENCES

[1] Y.A. Reznik et al., "User-adaptive mobile video streaming," in *Visual Communications and Image Processing*, 2012.

[2] R. Vanam and Y.A. Reznik, "Improving the efficiency of video coding by using perceptual preprocessing filter," in *Data Compression Conference*, 2013, p. 524.

[3] Y.A. Reznik and R. Vanam, "Improving coding and delivery of video by exploiting the oblique effect," in *Proc. 1st IEEE Global conference on signal and information processing*, Dec. 2013, pp. 775–778.

[4] 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*, Dec. 2013, pp. 426–429.

[5] Toshiyuki Fujine, Yasuhiro Yoshida, and Michiyuki Sugino, "The relationship between preferred luminance and tv screen size," in *Proc. SPIE*, 2008, vol. 6808, p. 68080Z.

[6] Alan C. Bovik, *Handbook of image and video processing*, AP, 2005.

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

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

[9] Scott 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.

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

[11] J.A. Movshon and L. Kiorpes, "Analysis of the development of spatial contrast sensitivity in monkey and human infants," *JOSA A*, vol. 5, no. 12, pp. 2166–2172, 1988.

[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] John G Nathan, Daniel R Anderson, Diane E Field, and Patricia Collins, "Television viewing at home: Distances and visual angles of children and adults," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 27, no. 4, pp. 467–476, 1985.