

# Perceptual pre-processing filter for user-adaptive coding and delivery of visual information

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

**Abstract**—We describe design of an adaptive video delivery system employing a perceptual preprocessing filter. Such filter receives parameters of the reproduction setup, such as viewing distance, pixel density, ambient illuminance, etc. It subsequently applies a contrast sensitivity model of human vision to remove spatial oscillations that are invisible under such conditions. By removing such oscillations the filter simplifies the video content, therefore leading to more efficient encoding without causing any visible alterations of the content. 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.

## I. INTRODUCTION

The eventual destination of visual information delivered to a reproduction device (TV, tablet, phone) is a viewer who is looking at it. However, not all information presented on the screen may be useful, as it may include harmonics that are not distinguishable by human vision under current viewing conditions. Factors that affect perception include size, brightness, and pixel density of the screen, distance between the viewer and the screen, ambient illuminance, etc.

In most conventional video coding and delivery systems such parameters are not known exactly, and only assumed to be within a certain range (e.g. viewing distance equal to 3–4 × height of the screen). However, as exemplified in Figure 1, it is conceivable to design an adaptive system that would measure such characteristics dynamically and pass them back to the transmitter. In turn, the transmitter may use this information for more effective encoding of visual information 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.

This paper discusses design of a pre-processing filter suitable for use in such a system<sup>1</sup>. Our design exploits two basic phenomena of human vision [3]–[5]:

- *contrast sensitivity function (CSF)* [5] - relationship between frequency and contrast sensitivity thresholds of human vision, and
- *eccentricity* - rapid decay of contrast sensitivity as angular distance from gaze point increases.

Both 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) [6], SQRI metric [5], S-CIELAB [7], etc. Previously suggested applications of eccentricity included coding with eye-tracking feedback, foveal coding [3], etc.

<sup>1</sup>For our previous work on this topic see [1], [2].



Fig. 1. Architecture of user-adaptive video delivery system employing perceptual pre-processing filter.

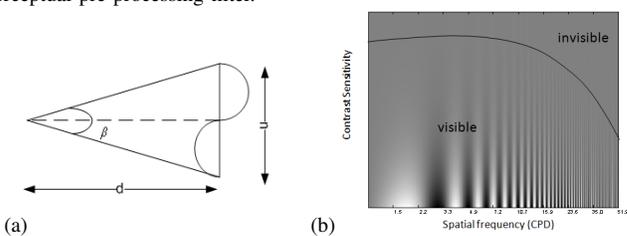


Fig. 2. (a) Illustration of the concept of spatial frequency. (b) Contrast sensitivity function (CSF) of human vision.

However, our application is different. We are not suggesting to 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. By removing such oscillations our filter simplifies video content, thereby leading to more efficient encoding without causing visible alterations of the content. 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. In our experiments 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.

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. Underlying principles

The key phenomena used by our filter is known as the Contrast Sensitivity Function (CSF) of human vision [5]. We illustrate it in Figure 2.b. As customary, spatial frequency is expressed in *cycles per degree* [cpd], and the contrast sensitivity is defined as the inverse of contrast thresholds:

$$C_T = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = \frac{\text{amplitude}(I)}{\text{mean}(I)} \quad (1)$$

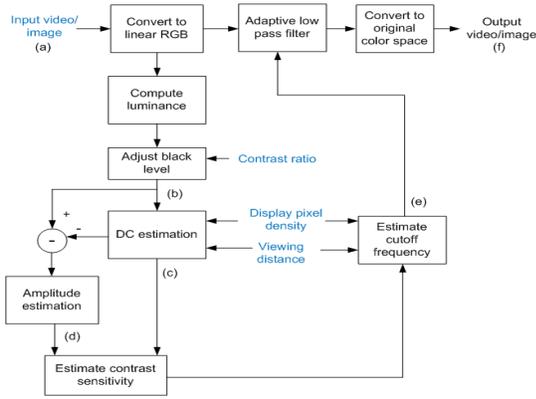


Fig. 3. Block diagram of our perceptual filter. The parenthesized letters refer to sub-figures in Figure 4.

where  $I_{\max}$ ,  $I_{\min}$  denote minimum and maximum intensities of an oscillation. As further exemplified in Figure 2.a, the spatial frequency  $f$  of a sinusoidal grating with cycle length of  $n$  pixels can be computed as:

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

where  $\rho$  is the display pixel density (expressed in ppi),  $d$  is the distance between viewer and the screen (in inches), and  $\beta$  is the angular span of one cycle of the grating (in degrees).

The frequency corresponding to a point on CSF curve reaching contrast sensitivity 1 is called *visual acuity limit*. This is the highest frequency visible by people with normal vision.

We must also note that CSF characteristic is meaningful only for characterizing sensitivity to features localized in some small (about 2 degrees of viewing angle) spatial regions. Larger regions cannot be examined with the same acuity due to the eccentricity of human vision. This gives us an important cue on how to apply the CSF in our filter design.

### B. Filter design

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 Baboon input image, shown in Figure 4(a), as an example.

*a) 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:

$$y' = \alpha + (1 - \alpha)y, \quad (3)$$

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

*b) DC estimation:* 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 4 (c) illustrates the local DC

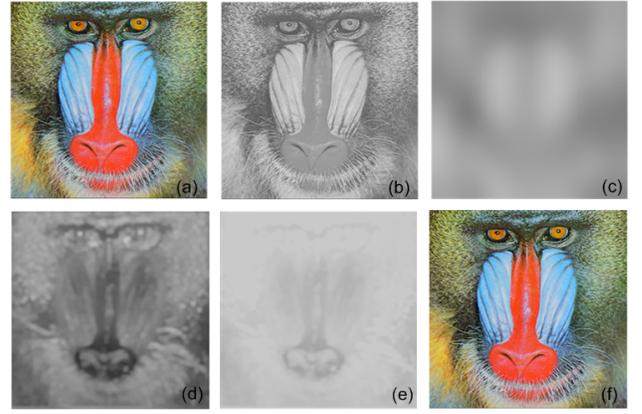


Fig. 4. (a) “Baboon” test image, (b) black level adjusted luminance, (c) local DC estimate, (d) amplitude envelope estimate, (e) cutoff frequency, and (f) filtered output image.

estimate after low pass filtering. We denote the DC value at location  $(i, j)$  as  $DC_{ij}$ .

*c) 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 4 (d) illustrates the amplitude envelope image. Let  $\text{amplitude}_{ij}$  be the amplitude at location  $(i, j)$ . The contrast sensitivity at location  $(i, j)$  is subsequently computed as

$$x_{ij} = \frac{DC_{ij}}{\text{amplitude}_{ij}}. \quad (4)$$

*d) Cutoff frequency estimation:* 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 CSF peaks, and  $F_{\max}$  is the visual acuity limit.

For instance, when employing the Movshon and Kiorpes CSF model [8], this yields the following algorithm for computing the highest visible frequency  $f_c(x_{ij})$ :

$$f'_c(x_{ij}) = -42.26 + 78.46x_{ij}^{-0.079} - 0.049x_{ij}^{1.08}$$

$$f_c(x_{ij}) = \begin{cases} F_{\min}, & f'_c(x_{ij}) < F_{\min} \\ f'_c(x_{ij}), & F_{\min} \leq f'_c(x_{ij}) \leq F_{\max} \\ F_{\max}, & f'_c(x_{ij}) > F_{\max}. \end{cases} \quad (5)$$

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

*e) Filtering operation:* Once cutoff frequency  $f_c(i, j)$  at location  $(i, j)$  is obtained, it is passed as a parameter to a low-pass filter. Equation (2) is used to map  $f_c(i, j)$  to pixel domain. We operate this filter in linear space, followed by conversion to the desired output color format. Figure 4 (h) illustrates the final filtered image.

## III. EXPERIMENTAL SETUP AND RESULTS

### A. Video sequences and encoder settings

In our experiments we used the following standard video test sequences: “IntoTrees,” “DucksTakeoff,” [9], “Life” [10],

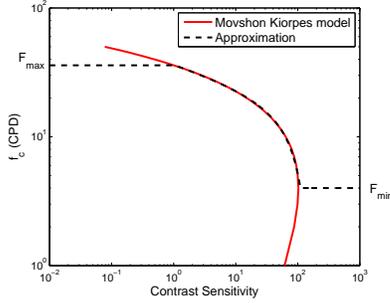


Fig. 5. Computing cutoff frequency using partially inverted CSF model [8].

and “Sunflower” [11]. We used x264 encoder [12], instructed to produce High-Profile H.264/AVC - compliant bistreams. To produce encodings of both the original and filtered content with similar distortion we instruct codec to apply same QPs in encodings of both sequences. Specific choices of QP values that we selected for each sequence are shown in Table I. These QPs were found to produce encodings of the original (non-filtered) sequences at approximately 10Mbps and 5Mbps rates, which we felt are practically relevant. For sequence “Sunflower” we only tested the 5 Mbps operating point, as it is substantially easier one to encode.

TABLE I  
SEQUENCES AND QPS USED @ 10 MBPS AND 5 MBPS.

Sequence	Resolution	10 Mbps		5 Mbps	
		QP	PSNR	QP	PSNR
IntoTrees	1080p, 25fps	27	35.7	30	34.3
DucksTakeOff	1080p, 25fps	38	28.2	42	26.1
Life	1080p, 30fps	25	39.4	29	36.8
Sunflower	1080p, 25fps	–	–	22	43.2

### B. Viewing conditions

Instead of tabulating results in terms of different viewing distances, we decided to use an angular characteristic: user’s observation angle that captures the width of the display, which we call the viewing angle  $\gamma$ . It is connected to the display width ( $w$ ) and viewing distance ( $d$ ) as follows:

$$\tan\left(\frac{\gamma}{2}\right) = \frac{w}{2d} = \frac{\text{width}[\text{pixels}]}{2\rho d} \quad (6)$$

This metric is convenient as results are applicable to different screen densities and sizes.

In our experiments we have set 12 operating points covering the range of observation angles from  $6^\circ$  to  $45^\circ$ . We have also tested the following contrast ratios:  $CR \in \{2:1, 3:1, 5:1, 10:1, 100:1, 1000:1, 10000:1\}$ . The first few correspond to situations when display is under sunlight, while last assumes studio monitor in dark room.

### C. Comparisons and verification

Given the above sets of operating points we have run our perceptual filter to produce sequences filtered for each combination of contrast and viewing angle parameters.

For comparison, we also produce sequences filtered using a conventional low-pass filter, with cut-off frequency selected

as follows:

$$f_c^{\text{uniform}} = f_c \left( \frac{1}{C_{\text{max}}} \right) \quad (7)$$

where

$$C_{\text{max}} = \frac{y_{\text{white}} - y_{\text{black}}}{y_{\text{white}} + y_{\text{black}}} = \frac{CR - 1}{CR + 1} \quad (8)$$

is the maximum contrast achievable by the display.

We performed two types of comparisons:

- size of encoded original vs perceptually filtered sequences; and
- size of encoded uniformly filtered vs perceptually filtered sequences.

The first comparison is indicative of absolute gains achievable by a system employing our perceptual filter vs. conventional encoding. The second comparison is indicative of the effect of performing locally adaptive perceptual filtering vs. uniform filtering of the entire image with a single global cut-off.

To ensure the same level of quality in encodings of original and filtered sequences we have used the same encoder settings and the same fixed QPs. In addition, we have also performed visual cross-checks with the goal of verifying that under specified conditions both encoded original and encoded filtered sequences look identical. Simultaneous double-stimuli viewing was performed by a panel of 5 viewers. We did this for 3 viewing angles ( $30^\circ$ ,  $20^\circ$ , and  $10^\circ$ ) and effective contrasts of 100:1, and 10:1, and found no noticeable differences.

### D. Results

We present results for each sequence in our tests in Figures 6-9. Left side plots in all figures show rate savings w.r.t. non-filtered encoding. Right side plots show gains of our proposed adaptive filter w.r.t. uniform filter applied under same viewing conditions.

As expected, narrower viewing angles lead to improved compression by using perceptual pre-filter. Lower contrast ratios also lead to some improvements, but this is noticeable only in the low-end of the range ( $CRs \leq 10:1$ ). We also notice that the amount of gain tends to be content-dependent. For example, at smallest viewing angle and lowest contrast point filtering of the sequence “IntoTree” results in over 70% gain, while for sequence “Sunflower” we only achieve  $\sim 40\%$ .

Comparison of uniform vs. our local contrast sensitivity-driven perceptual filter also shows significant content dependency. The biggest gains are observed for the sequence “IntoTree,” where our approach increases gain by about 35%, while for sequences such as “DucksTakeOff” and “Sunflower” the gains are only about 5-10%. We also notice that our proposed filter helps the most in a certain range of viewing angles and contrasts. Thus based on Figures 6-9 the gains are most significant when viewing angles are close to the range of  $12^\circ \dots 32^\circ$ , and when the contrasts are small ( $\leq 10:1$ ).

## IV. CONCLUSIONS

We have described the design of a pre-processing filter for a user- and environment-adaptive video delivery system. Such a filter receives parameters of the reproduction setup, such as viewing distance, pixel density, and ambient contrast of the

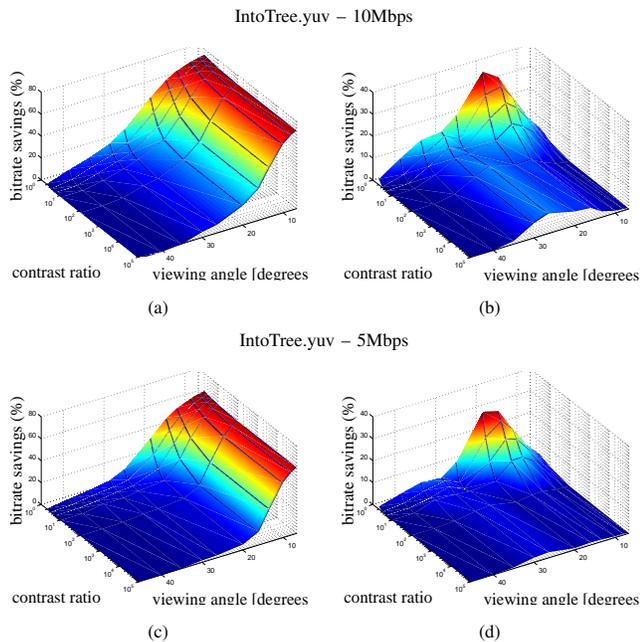


Fig. 6. Bitrate savings for: (a,c) perceptually filtered vs. non-filtered encodings, and (b,d) perceptually filtered vs. uniformly-filtered encodings of sequence “IntoTree”.

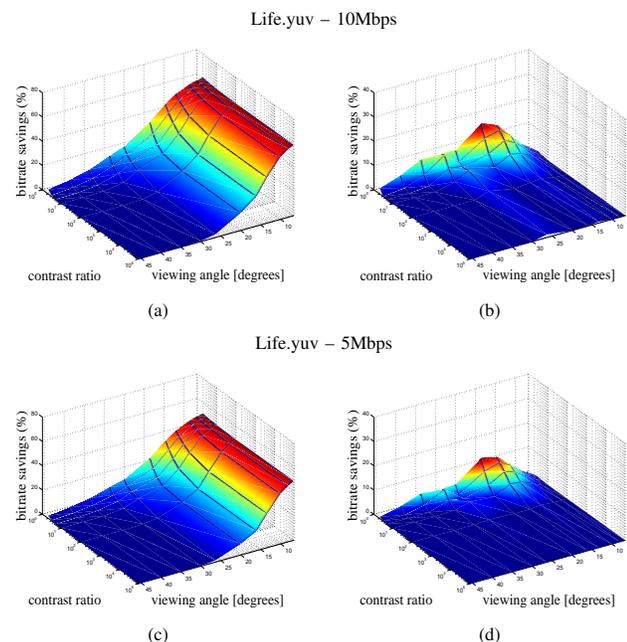


Fig. 7. Bitrate savings for: (a,c) perceptually filtered vs. non-filtered encodings, and (b,d) perceptually filtered vs. uniformly-filtered encodings of sequence “Life”.

display, and uses this information to remove spatial oscillations that are invisible under such conditions.

Through experiments, we have shown the use of our pre-filter may yield up to 70% bit rate savings compared to conventional encoding. We have also compared our filter with a conventional low-pass filter with an appropriately selected cut-off frequency, and have shown that it offers up to 35% reduction in the bit rate. Such improvements are particularly noticeable in low-contrast regimes.

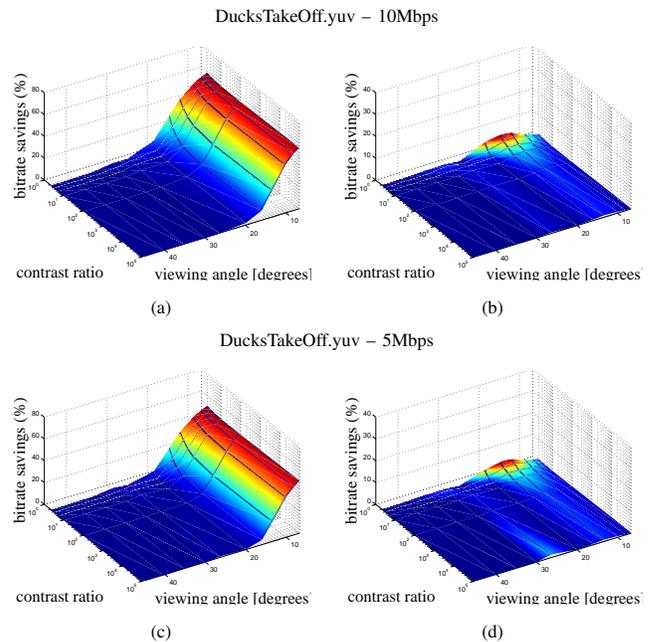


Fig. 8. Bitrate savings for: (a,c) perceptually filtered vs. non-filtered encodings, and (b,d) perceptually filtered vs. uniformly-filtered encodings of sequence “DucksTakeoff”.

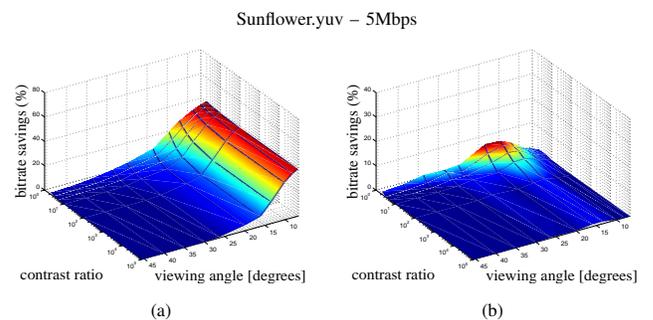


Fig. 9. Bitrate savings for: (a) perceptually filtered vs. non-filtered encodings, and (b) perceptually filtered vs. uniformly-filtered encodings of sequence “Sunflower”.

## REFERENCES

- [1] Y. Reznik *et al.*, “User-adaptive mobile video streaming,” in *Visual Communications and Image Processing*, 2012.
- [2] R. Vanam and Y. Reznik, “Improving the efficiency of video coding by using perceptual preprocessing filter,” in *Data Compression Conference*, p. 524, 2013.
- [3] A. C. Bovik, *Handbook of image and video processing*. AP, 2005.
- [4] H. Wu and K. Rao, *Digital Video Image Quality and Perceptual Coding*. CRC Press, 2005.
- [5] P. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality*. SPIE Press, 1999.
- [6] 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*, pp. 2–15, SPIE, 1992.
- [7] 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, pp. 731–734, SID, 1996.
- [8] J. 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.
- [9] “The SVT high definition multi format test set.” [ftp://vqeg.its.bldrdoc.gov/HDTV/SVT\\_MultiFormat/](ftp://vqeg.its.bldrdoc.gov/HDTV/SVT_MultiFormat/).
- [10] “Hdgreetings.” <http://www.hdgreetings.com/other/ecards-video/video-1080p.aspx>.
- [11] C. Keimel, J. Habigt, T. Habigt, M. Rothbuecher, and K. Diepold, “Visual quality of current coding technologies at high definition iptv bitrates,” in *IEEE MMSP*, p. 390–393, 2010.
- [12] “x264 encoder.” <http://www.videolan.org/developers/x264.html>.