Adapting Objective Video Quality Metrics to Ambient lighting

Louis Kerofsky, Rahul Vanam, Yuriy Reznik
InterDigital Communications Inc.
San Diego, CA USA
{louis.kerofsky, rahul.vanam, yuriy.reznik} @interdigital.com

Abstract— the limitations of objective metrics such as PSNR in evaluating video quality are well known to experts but less known to many users. A video tool which exploits perceptual phenomena may give improved subjective quality but lower objective performance when evaluated by such metrics. This presents a problem when describing the performance of perceptually motivated algorithms to general users. We propose and evaluate a method of extending existing objective metrics to account for perceptual factors such as viewing distance and ambient contrast. After describing the proposed algorithm we examine the variation with ambient lighting of the proposed modification. Subjective viewing tests are used to confirm the behavior of the extended objective metrics.

Keywords—video quality, objective metric, contrast sensitivity function

I. INTRODUCTION

Video quality metrics are in use to evaluate quality of video compression, delivery and display. An important application is providing a summary of the performance of a video compression algorithm. Metrics range greatly in degree of sophistication, from simple mean square error based comparison to a reference such as with PSNR, to sophisticated human visual system models such as Barten [1], Visual Difference Predictor [2], Sarnoff Just Noticeable Difference [3], and others. Similarly metrics differ in their assumptions about an available reference image. The metrics mentioned above are full-reference requiring a full reference image for definition. The assumption of a reference image can be limited as with reduced reference metrics or entirely eliminated with non-reference quality metrics. A typical application is evaluating the impact of compression artifacts in a video delivery system. Understanding the visibility of artifacts is essential to such applications.

The visibility of artifacts is determined both by the quality of the content sent to a display and by the viewing conditions of a viewer. Much work has been done on the quality metrics applied to the content to evaluate the severity of video artifacts. Less studied is the interaction with the viewer. Often specific viewing conditions are implicitly or explicitly specified. Thus one problem is to incorporate viewing conditions explicitly in the quality model calculations.

Limitations of PSNR as a visual quality metric have been well discussed in the technical community, for example in [4]. Objective metrics based on SSIM or MS-SSIM [5] have become popular recently, Despite the known limitations, PSNR is commonly used to evaluate system performance and application such as tuning encoding parameters due to the simplicity and application to a narrow range of parameter changes. In a limited application such as deciding between tools in a video codec and when used by an expert aware of its limitations, PSNR can be a valuable tool.

A problem arises when interacting with a less technical group. It is common to be asked about PSNR performance of a product when introducing it to the market place for instance. Academic references about the inadequacy of PSNR in capturing visual performance are of limited use when the customer demands a PSNR number. For example, it is well known that visual perception phenomena can be exploited to improve the performance of a compression system. Invisible but complex detail can be removed reducing the necessary bitrate to achieve similar subjective quality. As an example, the oblique effect, in which a viewer is less sensitive to diagonal frequency than to frequency in the cardinal directions, has been exploited to remove diagonal high frequency content to simplify encoding without impacting perceived quality [6], [7].

Such perceptually invisible modifications to an image will degrade a simple objective metric such as PSNR. Thus, focus on an objective metric such as PSNR may mask the benefits of perceptual processing. We are faced with the following problem: how to use the language of traditional PSNR familiar to the customer while communicating the performance benefits of system exploiting perceptual effects.

To address the problems described above, a method for incorporating viewing conditions, particularly viewing distance, with existing objective metric calculations was proposed in a recent publication [8]. This technique was applied to PSNR, SSIM, and MS-SSIM and compared to the results of subjective MOS tests performed over same content and parameters of viewing setup. It was shown that adapted versions of PSNR, SSIM, and MS-SSIM show similar behavior as MOS scores with changing viewing distance. In this paper we examine the adaptation to ambient viewing conditions with the viewing distance fixed.

The remaining sections of this paper are organized as follows. Section II provides a description of the modifications of traditional PSNR to account for perceptual factors of display contrast and viewing distance. Section III provides results of extended objective metrics computed for example sequences and viewing conditions. Section IV presents results from subjective testing which used the same sequences and conditions used for objective calculations. Section V provides our conclusions. An appendix is provided which gives some mathematical details.

II. ALGORITHM DETAILS

A. Fundamentals

It is well known that the perception of visual quality depends upon the viewing conditions. In psychophysics, the viewing distance is often expressed as a number of picture heights rather than physical units. The contrast sensitivity function (CSF) defines for a given viewing distance the minimum contrast needed for a spatial modulation to be visible. The spatial modulation is described by a frequency in cycles per visual angle. An example plot of a contrast sensitivity function is shown in Fig. 1 with indications of the visible and invisible regions of contrast for various spatial frequencies measured in cycles per degree of visual angle.

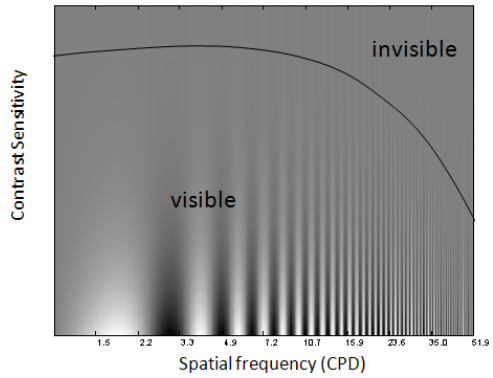

Fig. 1 Contrast Sensitivity Function.

Mathematical models of the CSF are commonly included in advanced visual models and quality metrics. A detailed mathematical model of the CSF is given in [12]. A summary of this model and its approximation and use are discussed in detail in the appendix.

B. Impact of Ambient light

The display is a vital component of a video reproduction system. The capability of the display and viewing condition i.e. ambient light and viewing distance greatly impact visibility of image artifacts. The contrast ratio defines a maximum contrast achievable for any image shown on the display. Ambient light has two primary impacts on the display system. The first impact is on the visual system of the viewer through the mechanism of adaptation. The second impact is the addition of light reflected from the display reducing the effective contrast ratio of the display and giving a washed-out effect.

In the first effect, the viewer adapts to the luminance of the surround of the display. Barten has modelled this as a scaling of the CSF function as the surround luminance differs from the target luminance [12]. This scaling factor is reproduced below in (1). Where the quantity L_S is the luminance of the surround, L is the luminance of the test object, and X_o is the size of the object in visual degrees. The basic CSF formula is multiplied by this scaling term when the ambient differs from the stimulus luminance. The CSF has greatest sensitivity when the surround luminance matches the target luminance.

$$f = e^{\frac{ln^2\left(\frac{L_S}{L}\left(1+\frac{144}{X_o^2}\right)^{0.25}\right)-ln^2\left(\left(1+\frac{144}{X_o^2}\right)^{0.25}\right)}{2\cdot ln^2(32)}} \quad (1)$$

In the second effect, ambient light reflected from the display surface reduces the effective display contrast giving a washed-out image appearance. The Michelson contrast of an image depends upon the maximum I_{max} and minimum values I_{min} as in (2).

$$C_{Michelson} = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \in [0,1] \quad (2)$$

In the presence of ambient illumination, a fraction of light is reflected from the display surface. Given an ambient illuminance of A lux and a display with reflectivity r, the luminance of the reflected image L_r in cd/m^2 is given by (3). This formula assumes diffuse (Lambertian) type of reflectance. This reflected light is added to the luminance of an image on the display as shown in (4). If the image values are relative for instance on a scale [0,1] an equivalent offset can be used by scaling the additive luminance by the maximum display luminance L_{max} as shown in (5), where $I_{original}$ is the original image, and I_{offset} is the resulting image after addition of the offset.

$$L_r = \frac{r \cdot A}{\pi} \quad (3)$$

$$L_{total} = L_{image} + L_r \quad (4)$$

$$I_{offset} = I_{original} + \frac{L_r}{L_{max}} \quad (5)$$

The impact of light reflected from the display is most significant on dark images. A simulation of this additive reflected light is seen by comparing Fig. 2 and Fig. 3. In this simulation a maximum luminance of 200 nits, a native contrast ratio of 500:1, and a display reflectivity of 10% were assumed. The loss of dark detail due to 1000 lux ambient compared to 10 lux ambient is visible in comparing the images.

Fig. 2 "Dark Forest" ambient 10 lux

Fig. 3 "Dark Forest" ambient 1000 lux

C. Cut-off frequency determination

Given a finite contrast ratio, the highest frequency which is visible under a given CSF and viewing distance can be determined. The limit on display contrast determines a lower bound on the contrast sensitivity achievable by the display.

Spatial frequencies above the cut-off frequency require contrast levels greater than that achievable on the display to be visible. Thus, given a display contrast ratio we can determine a cut-off frequency beyond which detail will not be visible. This concept is illustrated in Fig. 4. Pairs of spatial frequency and contrast sensitivity lie at various points on this plot. The region above the CSF curve is invisible to the viewer. The horizontal line of this figure indicates the minimum necessary sensitivity implied by an upper bound on display contrast. Pairs of frequency and sensitivity above this line are not representable on the display. The minimum sensitivity and the highest visible spatial frequency are linked by the CSF. Mathematically, the upper bound on the display contrast C determines a lower bound on the achievable contrast sensitivity S_{min}. The CSF relates a lower bound on sensitivity to an upper bound on visible frequency.

We use an approximation of the inverse CSF model for determining the cut-off frequency f_c from the maximum contrast C_{max} achievable on the display in a given ambient environment. Details on the inversion of the CSF model are supplied in the appendix.

D. *Visual frequency conversion*

The spatial frequency $f_c(s)$ given by the inverse CSF is in units of cycles per visual degree. For use in image processing these need to be converted to pixel specific units. The conversion between visual angle and pixels relies upon the viewing distance and display pixel density i.e. pixels-per-inch ppi. The relevant geometry is shown in Fig. 5. The spatial frequency, f, of a sinusoidal grating with cycle length n pixels can be computed as in (6). In this expression, the distance from viewer to display is d inches. The visual angle corresponding to this cycle is β degrees. The display density is p pixels per inch. When the viewing distance D is expressed in picture heights, the height of the display in pixels H_{pixels} is used in place of the density p in this expression to give (7).

$$f = \frac{1}{\beta}\,[cpd], \beta = 2\tan^{-1}\left(\frac{n}{2d\rho}\right) \tag{6}$$

$$f = \frac{1}{\beta}\,[cpd], \beta = 2\tan^{-1}\left(\frac{n}{2D \cdot H_{pixels}}\right) \tag{7}$$

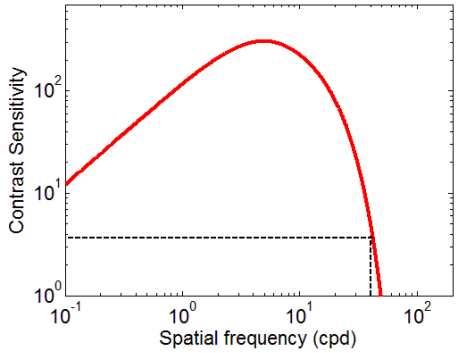

Fig. 4 Relating minimum sensitivity and maximum visible frequency via the CSF.

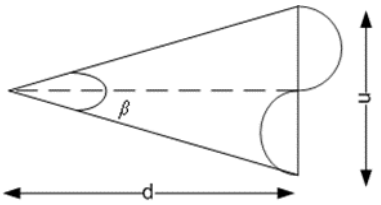

Fig. 5 Geometry relating visual angle, viewing distance and cycle length.

E. *Display Ambient application*

These models are used in a display application by linking various parameters to display and environmental properties. We model the surround as given by a typical surround reflectance of 18% and the display is taken as the object used in the Barten scaling function described in Section II.B. We consider three sample display devices: Phone, Tablet, and TV and evaluate the highest visible frequency under increasing ambient. Note these each have a different maximum screen brightness, native contrast ratio and screen reflectivity summarized in Table 1. These are chosen to represent a variety of sample values.

Table 1 Display parameters

	L_{max}	CR_{native}	*reflectivity*
Phone	100 nits	100:1	10%
Tablet	200 nits	300:1	6%
TV	450 nits	1000:1	4%

Based on these parameters and the models described above, we can compute the highest visible frequency under various ambient light level. This accounts for both the scaling of the CSF due to surround effect and the reduction in effective contrast ratio due to ambient reflection. A plot of the cut off frequency versus ambient light level is provided in Fig. 6

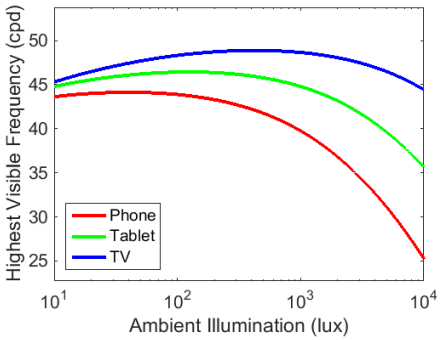

Fig. 6 Highest visible frequency vs. ambient.

F. *System Diagram*

A diagram of the algorithm defining the extension of an objective metric to include viewing condition dependence is shown in Fig. 7. We assume an objective metric $M(I_1,I_2)$ for computing a full reference image quality metric is given, for instance M could be PSNR, SSIM, or MS-SSIM. We assume display parameters are given. Specifically the display maximum luminance, the display reflectivity, the display

native contrast ratio, and the display pixel density (or equivalently height in pixels) are used. We assume the viewing conditions are described by two values, the ambient light level A and the viewing distance D.

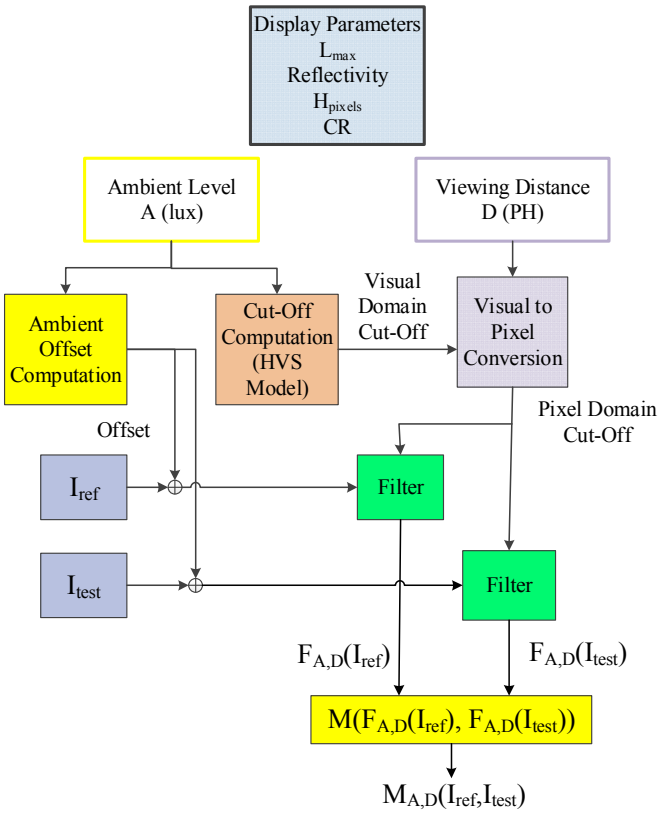

Fig. 7 System diagram extending metric *M*() to include viewing conditions in metric $M_{A,D}$().

The display maximum luminance, display reflectivity and ambient light level are used to compute an offset due to reflected light as defined in (5). The ambient light level and display properties are used with an HVS model to determine a maximum visible spatial frequency, cut-off frequency, expressed in cycles-per-(visual)-degree (cpd) as described in Section II.C. The viewing distance and display pixel density are used to convert this cut-off frequency into the pixel domain defining a pixel domain cut-off frequency in cycles-per-pixel (cpp) as in Section II.D. As shown in Fig. 7, both the reference and test images are processed in the same manner adding equal offsets prior to applying a low-pass filter based on the cut-off frequency. The reference image, I_{ref}, and the image under evaluation, I_{test}, are each filtered to produce modified images. The result of this processing becomes new inputs to the given objective metric. Thus the extended objective metric is the original objective metric evaluated on a pair of images which have been pre-processed based on display and viewing conditions.

For a given set of display parameters, for instance assuming typical values, the extended objective metric includes the additional parameters of ambient level, A lux, and viewing distance D in picture heights.

The proposed extended objective metric $M_{A,D}$ is defined as the given metric M evaluated on the processed reference and test images. This process is summarized in (8).

$$M_{A,D}(I_{ref}, I_{test}) = M\left(F_{A,D}(I_{ref}), F_{A,D}(I_{test})\right) \quad (8)$$

In the above formula, $F_{A,D}(I)$ is the result of processing an image by two steps: adding an offset determined by the ambient A, and applying a low-pass filter determined by A and the viewing distance D. The values A and D represent the ambient level and viewing distance, respectively.

III. Objective Evaluation Including Ambient

We demonstrate example calculation of P-PSNR on sample content and varying viewing conditions. A fixed viewing distance is chosen and the effective contrast is changed by varying ambient lighting.

A. Methodology

Source material was generated by using two well-known full HD, 1920x1080, sequences from the video coding community "parkscene" and "sintel_trailer" [13]. The sequence "sintel_trailer" has low luminance in many sections which we expect to be most sensitive to ambient illumination. Each was encoded using the open source x264 encoder [9] with three fixed quantization levels QP = 26, 32, 38 giving a range of compression quality and artifacts. Each sequence was decoded to provide six sample video sequences, two content at three encoding levels each. For this evaluation, display parameters of the subjective test were used: L_{max}=185 cd/m^2, r=8%, CR=500:1, and screen height in pixels=1080. The viewing distance was set to 3H and surround reflectance to 18%.

For each ambient condition, an appropriate cut-off frequency and additive offset were selected based on a CSF model ambient light level and assumed display parameters. Three objective metrics PSNR, SSIM, and MS-SSIM were computed at each ambient light level between original and decompressed images after each was modified by additive offset and low-pass filter operations to give the perceptual extensions P-PSNR, P-SSIM and P-MSSSIM, respectively.

B. Results

The extensions of PSNR, SSIM and MS-SSIM metrics are plotted as a function of ambient light level for three different compressed versions of the sequences "parkscene" and "sintel_trailer" in Fig 8. Each plot uses a single sequence and metric while the ambient light level is varied through 10 lux, 100 lux, 1000 lux and 2000 lux for each coding conditions. As the ambient level increases the effective display contrast reduces making artifacts less significant. The result is an increase in objective metric values. The PSNR rises without bound as the ambient increases corresponding to the convergence between the reference and test images. As the amount of reflected light exceeds the display brightness, the display output would become the same constant white in all cases giving an infinite PSNR. The P-SSIM and P-MSSSIM metrics increase toward saturation as the ambient parameter increases. This saturation of quality is more representative of the expected property. The lower quality versions increase more rapidly than the higher quality versions of the sequence.

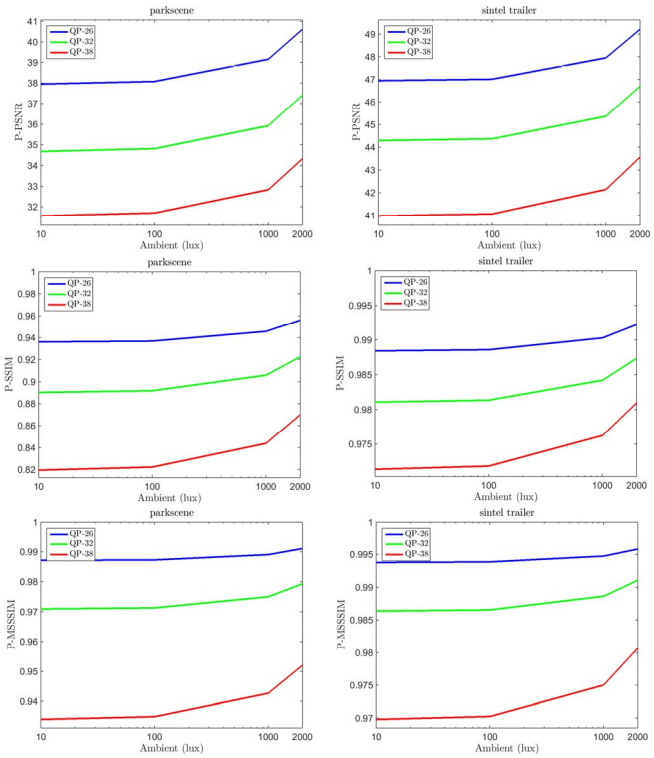

Fig. 8 Extended objective metrics (P-PSNR, P-SSIM, P-MSSSIM) versus ambient for three compression levels and two sequences.

IV. SUBJECTIVE EVALUATION

We investigate the subjective quality of fixed compressed video sequences as the ambient light level is varied in a viewing test. The sequences "parkscene" and "sintel_trailer" were the same as used in the objective test results of Section III. The sequences were each coded at three quantization levels QP = 26, 32, and 38 as with the objective metric tests. Seven viewers were used to rate the subjective quality of the video under different ambient conditions: 10 lux, 100 lux, 1000 lux and 2000 lux. With two sequences, three compression levels, and four ambient levels, each subject evaluated 24 trials.

A. Methodology

For testing, each of the compressed sequences was used without processing. Subjects were asked to rate the quality of video using the five point scale given in Table 2, as recommended by BT.500 [11]. Prior to scoring, subjects where shown a video sequence encoded with low quantization QP=20 and told this was an example of "Excellent". The same video sequence encoded with high quantization QP=46 was shown as an example of "Bad".

During evaluation subjects were placed at a viewing distances 3H from the display and the ambient light level was varied from low-to-high in a series of tests. At each ambient level, the subjects were shown the six compressed sequences in a random order and asked to provide a score. Sony LMD-941W reference monitor was used in our tests. Relevant display parameters were: L_{max}=185 cd/m^2, r=8%, CR=500:1, and the screen height in pixels=1080.

Table 2 Quality Scale

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

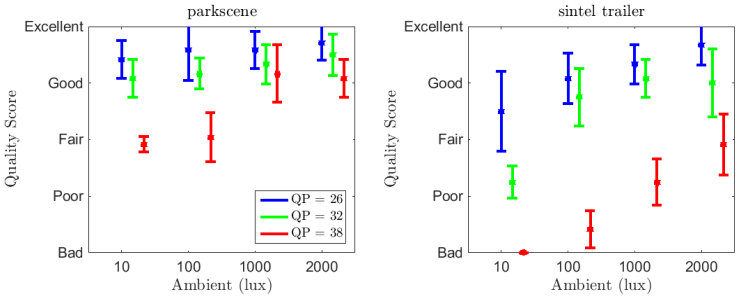

Fig. 9 Subjective test results under various ambient levels.

B. Results

Results of the subjective voting are shown in Fig. 9. In each plot, results from testing a single sequence encoded at three different quantization levels under a range of ambient light strengths is given. We notice that sequences rated as "Bad" or "Poor" under low ambient light improve in quality score as the ambient light level increases. This agrees with the observation that compression artifacts become less visible under elevated ambient light as well as the trends expressed in the modified objective metrics. The sequences rated at high quality "Good" or "Excellent" under low ambient light levels are largely independent of ambient light level i.e. compression artifacts were not seen under any conditions. The "sintel_trailer" sequence which showed the lowest quality under low ambient gave the highest increase in quality score as the ambient level rose.

The correlation between the subjective scores and each extended objective metric across ambient light levels was calculated for each sequence and QP value. Results are shown in Fig. 10. The correlations are nearly independent of QP but show some dependence on the sequence. With the exception of the P-PSNR on the "'sintel trailer" sequence the correlations are above 85%. The P-PSNR achieves only a 65% correlation with the subjective tests for the "sintel trailer" sequence.

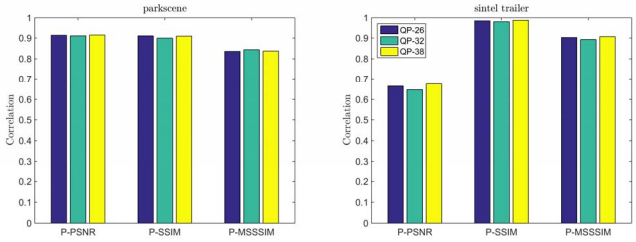

Fig. 10 Correlation between extended metrics and subjective scores.

V. Conclusions

We demonstrated a method for incorporating viewing conditions within a full reference objective video quality metric. Both the reference and the video under test are modified identical offset and filter operations determined by the ambient viewing conditions and display parameters. A low pass-filter accounts for reduced sensitivity to high frequency details, an offset accounts for ambient light reflected from the display. At present this is done prior to the metric calculation though it could be combined to define a new metric.

We have shown the effectiveness of the proposed method by producing perceptually-adapted versions of PSNR, SSIM, and MS-SSIM in the presence of compression artifacts due to quantization in a video codec by comparing their outputs to MOS scores produced by human observers. The variations in scores with viewing conditions factored in are absent in existing objective metrics. Unification of the prior distance based adaptation [8] and the ambient based modification of this work look straightforward and should be investigated. The methods presented work on full-reference metrics. The possibility to extend this approach to non-reference metrics is another possibility for future work.

VI. References

[1] P. Barten, "Contrast sensitivity of the human eye and its effects on image quality." Vol. 72. SPIE press, 1999.

[2] S. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in Proc. *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics, 1992.

[3] J. Lubin and D. Fibush, "Sarnoff JND Vision Model," T1A1.5 Working Group Document #97-621, ANSI T1 Standards Committee, 1997.

[4] Z. Wang and A. Bovik. "Mean squared error: love it or leave it? A new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98-117, Jan. 2009.

[5] Z. Wang, A. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process*ing, vol. 13, pp. 600–612, Apr. 2004.

[6] Y. Reznik and R. Vanam. "Improving the coding and delivery of video by exploiting the oblique effect," in Proc. *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 775-778, 2013.

[7] R. Vanam and Y. A. Reznik, "Improving the efficiency of video coding by using perceptual preprocessing filter," in Proc. *IEEE Data Compression Conference (DCC)*, pp. 524, 2013.

[8] L. Kerofsky, R. Vanam, and Y. Reznik, "Perceptual adaptation of Objective video quality metrics," In Proc. *Ninth International Workshop on Video Processing and Quality Metrics (VPQM)*, 2015.

[9] VideoLAN x264 encoder project described at http://www.videolan.org/developers/x264.html.

[10] Multimedia Signal Processing Group MMSPG at EPFL http://mmspg.epfl.ch/vqmt

[11] Rec. ITU-R BT.500-13, Methodology for the subjective assessment of the quality of television pictures, 2012.

[12] P. Barten, "Formula for the contrast sensitivity of the human eye." Electronic Imaging 2004. International Society for Optics and Photonics, 2003.

[13] HD videos, URL: https://media.xiph.org/video/derf/

Appendix: Calculation of cut-off frequency

A practical display has a limit on the achievable contrast under particular ambient viewing conditions. A CSF model relates this minimum contrast sensitivity to a maximum visible spatial frequency defined as the highest frequency where the CSF exceeds this minimum sensitivity.

The cut-off frequency is computed as follows: determine the maximum contrast achievable on the display on the display under existing conditions, C_{max}. The minimum sensitivity $S_{min} = 1/ C_{max}$. Find the solution of $S(u) = S_{min}$. This defines the cut-off frequency.

To determine the frequency at which S_{min} is achieved, a mathematical model of the human contrast sensitivity function developed by Barten [12] is summarized below. The sensitivity threshold of a spatial frequency of u cycles per degree is given by:

$$S(u) = \frac{Ae^{-Du^2}}{\sqrt{(B+u^2)\left(C+\frac{1}{1-e^{-0.002u^2}}\right)}}$$

where constants A, B, C, and D are defined below in terms of the object luminance L and the object size X_o.

$$A = \frac{5200E}{\sqrt{0.64}} \qquad B = \frac{1}{0.64}\left(1+\frac{144}{X_0^2}\right)$$

$$C = \frac{63}{L^{0.83}} \qquad D = 0.0016\left(1+{}^{100}/_{L}\right)^{0.08}$$

$$E = exp\left(-\frac{ln^2\left(\frac{L_s}{L_o}\left(1+\frac{144}{X_o^2}\right)^{0.25}\right) - ln^2\left(\left(1+\frac{144}{X_o^2}\right)^{0.25}\right)}{2\,ln^2(32)}\right)$$

For large u, this can be approximated by

$$S_1(u) = \frac{Ae^{-Du^2}}{\sqrt{(B+u^2)(C+1)}}$$

This model is a function of the viewing conditions. For given viewing conditions, the display brightness and size determine the constants L and X_0. Thus the function $S_1(u)$ is determined.

The function $S_1(u)$ can be analytically inverted to give:

$$u = S_1^{-1}(s) = \sqrt{\frac{LambertW\left(\frac{2DA^2e^2DB}{(C+1)s^2}\right)}{2D}}$$

Where $LambertW(z)$ is a solution of equation:

$$LambertW(z).e^{LambertW(z)} = z$$

The minimum sensitivity achievable on the display is related to the highest visible frequency through the CSF model.