Mobile Display Power Reduction for Video Using Standardized Metadata

Felix C. Fernandes, Esmaeil Faramarzi, Xin Li, Zhan Ma, *Member, IEEE*, and Xavier Ducloux

Abstract—The inordinate power consumption of contemporary display panels accelerates battery depletion in mobile devices. Existing display-power reduction approaches are suboptimal because they incur additional computations and latency on the mobile device or because they assume a server-client model in which a single entity controls the server and the client. To overcome these limitations, we extract *standardized* metadata at a server and transmit it to a client along with the video bitstream. This metadata includes quality indicators, associated power controls, and contrast-enhancement information. The metadata guarantees that specified quality levels are achieved while avoiding flicker and minimizing power consumption. Furthermore, the metadata has minimal computational, storage & power overheads and, because it is standardized, the metadata benefits servers and clients in open ecosystems wherein all entities function independently. Our research has been implemented on mobile devices and provides, on average, 32.5 percent power reduction at high quality and up to 84 percent power reduction at acceptable quality. The ISO/IEC Moving Pictures Expert Group has recently standardized our proposed metadata in the Green Metadata Standard.

Index Terms—Energy-efficient mobile device, display power reduction, standard, meta-data, backlight scaling

The unprecedented popularity of video applications has created a huge demand for increased battery-charge life on mobile phones and tablets. Therefore, the power consumption of video-processing elements in such devices must be reduced. Among these elements, the display subsystem consumes a large fraction of the power on mobile devices. Hence, to prolong battery-charge life on such devices, it is crucial to reduce display power consumption. In this article, we show how to reduce a device's power consumption by up to 84 percent with metadata that accompanies the video content being displayed. We obtain this metadata at a server by processing video-frame statistics temporally and then transmitting the metadata along with the associated video bitstream. The metadata has been standardized by the ISO/IEC Moving Pictures Experts Group (MPEG) in the Green Metadata Standard (ISO/IEC 23001-11) [1]. The metadata is transported in the ISO Base Media File Format (ISO/IEC 23001-10) [2] and in the MPEG-2 Transport Stream (ISO/IEC 13818-1) [3]. Fernandes et al. [4] provide a high-level overview of *all* power reduction technologies in the standard. This paper is restricted to display-power reduction and, unlike Reference [4], provides a comprehensive, detailed investigation of the subject.

- F.C. Fernandes is with Apple Inc., 1 Apple Park Way, Cupertino, CA 95014. E-mail: felixfcaf@gmail.com.
- E. Faramarzi is with Samsung, 1301 E. Lookout Dr, Richardson, TX 75082. E-mail: efaramarzi@gmail.com.
- X. Li is with the Startup Company, 753 Player Dr, Plano, TX 75025. E-mail: neasoon@gmail.com.
- Z. Ma is with Nanjing University, 22 Hankou Rd, Gulou Qu, Nanjing Shi, Jiangsu Sheng, China 210008. E-mail: zhan.ma@gmail.com.
- X. Ducloux is with Harmonic Inc., 57 rue Clément Ader, Cesson Sévigné, France 35510. E-mail: xavier.ducloux@harmonicinc.com.

Manuscript received 27 Dec. 2016; revised 31 Oct. 2017; accepted 17 Apr. 2018. Date of publication 20 Apr. 2018; date of current version 3 Dec. 2018. (Corresponding author: Felix C. Fernandes and Zhan Ma.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2018.2829185

1 Related Work

The Perceived Luminance (PL) from a Liquid Crystal Display (LCD) is proportional to the input RGB subpixel values and to the backlight intensity of the display. Decreasing the backlight intensity by a Backlight Scaling Factor (BSF) reduces power and PL. To restore the PL, the RGB subpixel values are scaled up by a Subpixel Scaling Factor (SSF). For 8-bit RGB values, the SSF scaling will cause clipping at the maximum value 255 and some quality loss may occur if too many RGB subpixels are clipped. Image enhancement techniques can be used to mitigate the quality loss that would occur at large power-reduction levels. The research on DA has focused on either maximizing image quality at a given power-reduction level [5] or on power minimization for a given image quality level [6]. Although most techniques have focused exclusively on static images, some researchers have applied DA to *video content* [7], [8] and demonstrated video streaming with real-time DA implementations.

The preceding approaches assume that video is streamed to a mobile device which then analyzes the content to determine DA parameters. These parameters are applied to the mobile display to reduce power and to the video content to maintain quality. All DA-related processing occurs exclusively on the mobile client. However, Pasricha et al. [9] recognized that a server-client setting reduces power significantly by computing the DA parameters at the server and then transmitting them to the mobile client. This enabled the DA parameters to be computed optimally for entire video scenes so that flicker is mitigated at scene changes. Also, since the DA parameters are computed at a server, the mobile client does not have to waste battery power on the DA parameter

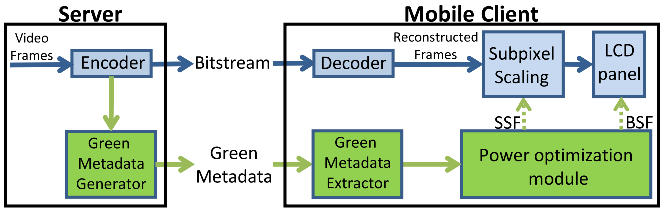

Fig. 1. The server-client model for an open ecosystem.

computation. Subsequently, Cheng et al. [10] demonstrated that offline processing in the server-client model mitigates flicker by smoothing backlight changes over successive video frames. Recently, Lin et al. [11], [12] deployed a cloud-based service supplemented by energy-saving mobile applications. The cloud server derives BSFs for popular videos on major streaming websites such as YouTube. These ratios are associated with the popular videos and stored on the cloud server. When a user starts to stream a popular video on a mobile device with the energy-saving mobile app installed, the app will retrieve the BSFs from the cloud server and apply DA during the streaming to save power. Note that the mobile app does not scale RGB values to restore quality. It merely synchronizes the backlight changes with the video playback so that display power consumption is reduced.

1.1 The Benefits of a Standards-Based DA Solution

In this article, we describe our research that extends the server-client DA approach [9], [10], [11], [12] by providing a standards-based solution that enables an ecosystem for the commercial deployment of DA. Fig. 1 shows the server-client model for an open ecosystem that uses Green Metadata. At the server, video frames are encoded to create a bitstream. A green-metadata generator analyzes the reconstructed frames and determines DA parameters that constitute the Green Metadata. Note that the reconstructed frames are stored within contemporary video encoders, and are identical to the decoded frames that will be displayed at the client. The bitstream and the Green Metadata are then transmitted to the client which decodes the bitstream and feeds the Green Metadata to a power optimization module. This module determines the BSF and the SSF from the Green Metadata, display properties, pre-configured user settings and from the remaining battery life of the mobile client.

Here are the main contributions of our research in which DA parameters are standardized as metadata:

1) The existing non server-client DA approaches [13] are prone to flicker at scene changes as demonstrated by Cheng et al. [10]. Because intensive pixel-wise processing is performed on a mobile device, these approaches also waste battery power, as observed by Hsiu et al. [12]. The existing server-client DA approaches [9], [10], [11], [12] work well in *closed ecosystems* in which a single entity controls both the server and the client. Unfortunately, most contemporary video streaming ecosystems are *open ecosystems* in which different entities control the server and the client. However, by using standardized metadata, the preceding drawbacks are overcome: (1) flicker is avoided by temporally smoothing the metadata, (2) mobile battery power is not wasted, and (3) metadata can be generated and applied by different entities in open ecosystems.
2) The DA parameters must be synchronized with associated video frames so that the display's BSF control is consistent with the displayed video's SSF adjustment. Accordingly, we have devised transport mechanisms in the ISO Base Media File Format (ISOBMFF) and in the MPEG-2 Transport System, the two most widely used systems-level protocols for open ecosystems. In Section 4, we describe usage of these protocols for DA.
3) The MPEG Dynamic Adaptive Streaming over HTTP (DASH) standard enables internet delivery of multimedia in open ecosystems. In Section 6, we explain how standardized DA metadata adds a new dimension to dynamic adaptive streaming: besides bandwidth considerations, media selection can be based on the media's power-consumption impact on the mobile display.
4) Several researchers have studied *image* contrast enhancement after DA [5], [6], [13], [14]. Only Raman et al. [15] have examined *video* contrast enhancement after DA. However, since they use exponential backlight smoothing to prevent flicker, the distortion constraint [11], [12] is violated. This constraint sets a lower bound on the BSF that may be assigned to an image frame, based on a pre-determined quality metric. In this article, we present a recursive smoothing algorithm that can be applied to both BSFs and to contrast bounds, while maintaining the distortion constraint.
5) As explained in Sections 2.3 and 2.4, our approach provides mobile clients with a list of Operating Points (OPs) ranging from maximum quality (lossless) to the minimum acceptable quality after applying contrast enhancement. The maximum quality OP enables the smallest power reduction and the minimum quality OP allows the largest power reduction. Thus our approach allows a mobile client to tradeoff between quality and power reduction based on its battery status. Although another server-client DA approach [11] does also tradeoff quality with power reduction, it does not scale RGB pixels or use contrast enhancement to maximize quality at the minimum quality OP.

To summarize, DA metadata had not been standardized before and that differentiates this work from others. Standardized metadata guarantees that specified quality levels are achieved while avoiding flicker and minimizing display power consumption in open ecosytems. Unlike other approaches, standardized metadata is easily synchronized with the associated video bitstream, enables power-adaptive streaming and incurs minimal computational, storage and power overheads.

2 Metadata-Assisted Display Adaptation

2.1 What is Display Adaptation?

The perceived luminance from a Liquid Crystal Display (LCD) is proportional to the input color-pixel values and to the backlight intensity of the display. Color pixels are typically composed of Red (R), Green (G) and Blue (B) subpixels. Let us denote a normalized subpixel value at Row i and Column j by $x(i,j)$. For 8-bit R, G and B subpixels, the normalization factor is 255. Then, the default luminance $\mathcal{L}_0(i,j)$ of a normalized subpixel $x(i,j)$ is the product of the subpixel transmittance $T(x(i,j))$ and the display's maximum backlight intensity B_M [13]

$$\mathcal{L}_0(i,j) = B_M T(x(i,j)). \tag{1}$$

The luminance may be decreased by applying S, a Backlight Scaling Factor (BSF), to scale down the backlight from its maximum value B_M. The BSF satisfies $S \in [0,1]$, and an example value of 0.1 indicates that the backlight intensity is at 10 percent of the maximum setting. Denoting the decreased luminance by $\mathcal{L}(i,j)$, we have

$$\mathcal{L}(i,j) = SB_M T(x(i,j)). \tag{2}$$

Because the relationship between display luminance and perceived brightness is non-linear, practical displays incorporate a gamma correction into the transmittance function T so that it implements the following non-linear mapping from [0, 1] to [0, 1]

$$T(x(i,j)) = x^{\frac{1}{\gamma}}(i,j)^1, \tag{3}$$

where γ denotes the gamma value of the display, and $\gamma < 1$.

Combining the preceding two equations, we get

$$\mathcal{L}(i,j) = (SB_M)x^{\frac{1}{\gamma}}(i,j). \tag{4}$$

Now, since S, the BSF, controls the maximum backlight voltage and display power consumption is proportional to the square of that voltage, we can reduce power consumption by reducing S. However, from Equation (4), decreasing S will lower the luminance $\mathcal{L}(i,j)$ and the viewer will perceive reduced brightness at the subpixel. Fortunately, Equation (5) shows that the luminance and consequently the perceived brightness can be maintained by simultaneously scaling the subpixel value $x(i,j)$ to $x(i,j)S^{-\gamma}$, as S is reduced

$$\mathcal{L}(i,j) = SB_M[x(i,j)S^{-\gamma}]^{\frac{1}{\gamma}}. \tag{5}$$

We shall refer to $S^{-\gamma}$ as the Subpixel Scaling Factor (SSF). Since the power consumption changes negligibly with subpixel variations, we can reduce the display power by dimming the power-consuming backlight by S, the BSF, and simultaneously scaling RGB values by $S^{-\gamma}$, the SSF. This power-reduction technique is called backlight dimming [14]. For an Organic Light-Emitting Diode (OLED) display, the luminance is proportional to the RGB values and to the supply voltage. Therefore, dynamic supply voltage scaling, which is conceptually similar to backlight dimming, can be used for power reduction in OLED displays [8]. In this article, we shall use the term Display Adaptation (DA) to refer generally to either backlight dimming or to dynamic supply voltage scaling. Unlike other approaches, we determine the BSF and the SSF from *standardized* metadata and from other mobile-client parameters. In subsequent sections, although x and $\mathcal{L}$ represent values corresponding to individual subpixels, we shall omit the subpixel coordinates (i,j) for notational simplicity.

1. For consistency with other works [13] in this field, we use $\frac{1}{\gamma}$ with $\gamma < 1$.

When the standardization process commenced, MPEG listed the following requirements [16] for the DA Green Metadata:

1) The metadata shall facilitate appropriate power consumption without the loss of Quality of Experience (QoE).
2) The metadata shall offer multilevel QoE and the means to choose between energy consumption and QoE.

Sections 2.2 and 2.3 address Requirements 1 and 2.

2.2 Metadata for Power Reduction without QoE Loss

Display panel inputs are designed to accept the full range of a specified bit depth. For example, an 8-bit display panel clips subpixel inputs to the range [0,255]. For normalized subpixel values, the display-panel input is clipped to the range [0,1]. Therefore, we assume that the domain of the transmissivity function $T(\cdot)$ is [0,1]. QoE loss occurs when scaled display panel inputs exceed the domain of $T(\cdot)$ and must be clipped. If the mobile client's battery level is high and if pre-configured user settings indicate a preference for operation without QoE loss, then such clipping must be avoided as described below.

From Equation (5), no QoE loss occurs if

$$xS^{-\gamma} \le 1. \tag{6}$$

Observe that

$$S^{-\gamma} \ge 1, \tag{7}$$

since $\gamma \in [0,1)$ and $S \in [0,1]$. Now, let x_{max} denote the largest subpixel in a reconstructed frame. If $x_{max} = 1$, then $S^{-\gamma} = 1$, from Equations (6) and (7). Therefore, $S = 1$ and no power saving is possible for $x_{max} = 1$. However, if QoE loss is tolerable, then power saving is possible when $x_{max} = 1$. This operating mode will be explained in Section 2.3.

If $x_{max} < 1$, then Equation (6) implies that $S \ge x_{max}^{\frac{1}{\gamma}}$. Setting $S = x_{max}^{\frac{1}{\gamma}}$ minimizes power consumption and from Equation (5), subpixel scaling by the SSF $S^{-\gamma}$ maps the display-panel input interval from $[0, x_{max}]$ to $[0, x_{max}S^{-\gamma}] = [0,1]$. Since the scaled subpixels are within the domain of the transmissivity function, no clipping occurs and we achieve power saving without any QoE loss.

To perform the backlight scaling by the BSF, $S = x_{max}^{\frac{1}{\gamma}}$, the client must know x_{max}. Clearly, the client can determine x_{max} for each decoded, reconstructed frame. However, as explained in Section 3, there are several advantages to generating and pre-processing video metadata at the server. Therefore, the statistic x_{max} is sent from the server to the client as Green Metadata to enable display power reduction without any QoE loss.

Fig. 2a shows the histogram of the normalized R, G and B subpixel values pooled together from a single reconstructed frame. Fig. 2b shows the histogram after subpixel scaling by the SSF which maps $[0, x_{max}]$ to $[0,1]$. Fig. 2c shows the output histogram after applying the SSF and the BSF to the display panel, as described in Equation (5). Observe that the output

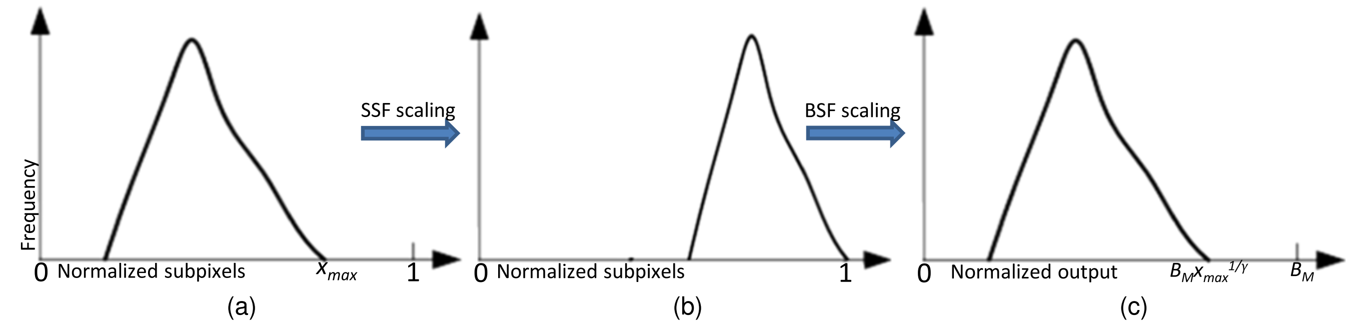

Fig. 2. (a) Histogram of normalized subpixels. (b) Histogram after scaling by the SSF. (c) Output histogram after applying SSF and BSF.

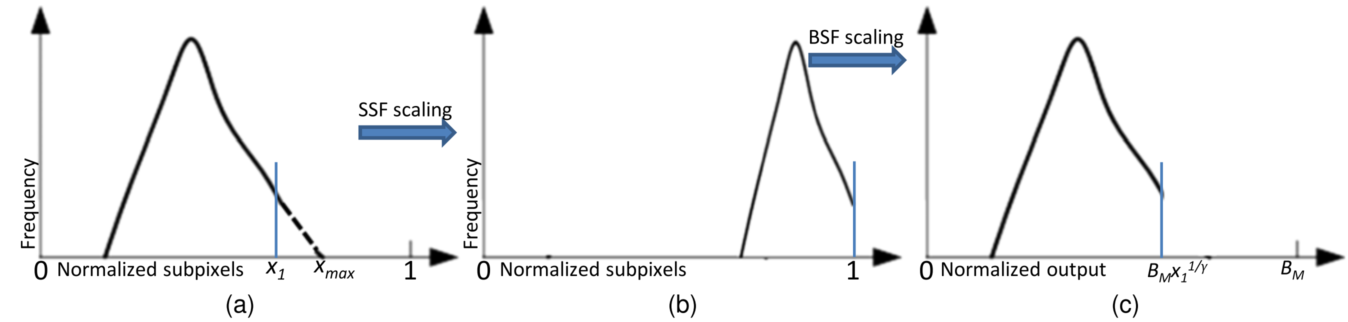

Fig. 3. (a) Histogram after clipping subpixels to x_1. (b) Histogram after scaling by the SSF. (c) Output histogram after applying SSF and BSF.

histogram is identical to that which would be obtained without backlight dimming, from Equation (1). Clearly, we can maintain QoE and reduce power consumption if $x_{max} < 1$.

2.3 Metadata for Multilevel QoE

To address Requirement 2 on multilevel QoE in Section 2, consider the histogram of a reconstructed frame shown in Fig. 2a. Using $S = x_{max}^{\frac{1}{\gamma}}$, as defined in Section 2.2, enables maximum power reduction without QoE loss because no subpixels are clipped when the SSF scales subpixels in Equation (5) which applies the SSF. However, we can achieve greater power reduction by decreasing the BSF below $S = x_{max}^{\frac{1}{\gamma}}$.

Specifically, if the BSF is set to $S = x_1^{\frac{1}{\gamma}}$, where $x_1 < x_{max}$, then power is reduced further because $x_1^{\frac{1}{\gamma}} < x_{max}^{\frac{1}{\gamma}}$. Unfortunately, Equation (6) is now violated because $xS^{-\gamma} = x(x_1^{\frac{1}{\gamma}})^{-\gamma} > 1$, for $x > x_1$. However, by clipping to x_1 all subpixels that are larger than x_1, then Equation (6) is no longer violated.

Fig. 3a shows the histogram of the reconstructed frames after clipping subpixels to x_1. Now, subpixel scaling by the SSF $S^{-\gamma}$, with $S = x_1^{\frac{1}{\gamma}}$ maps the subpixels in $[0, x_1]$ to $[0, 1]$ resulting in the scaled histogram in Fig. 3b. Finally, applying the BSF to the scaled subpixels results in a displayed output with the histogram shown in Fig. 3c.

2.3.1 Quantifying the QoE

Comparing Figs. 2c and 3c, we observe that some QoE has been sacrificed because contrast in the range $[B_M x_1^{-\gamma}, B_M x_{max}^{-\gamma}]$ has been lost. To quantify this QoE loss between the unscaled subpixels and the subpixels scaled by SSF and BSF in Fig. 3c, we use the Peak Signal to Noise Ratio (PSNR) in dB. For the PSNR computation, we revert to non-normalized subpixel values so that the largest subpixel value is denoted by $P = (2^N - 1)$ for N-bit color depth video. Thus, we obtain $\mathrm{PSNR}(x_1)$ by rounding Expression (8) and then clipping it to N bits

$$10 \log_{10}\left(\frac{P^2 \times W \times H \times N_{\text{color}} \times N_{\text{frames}}}{\sum_{n=1}^{N_{\text{frames}}} \sum_{c=1}^{N_{\text{color}}} \sum_{l=x_1+1}^{P} N_{c,n}(l) \times (l - x_1)^2}\right), \quad (8)$$

where, W and H are the video-frame dimensions, N_{color} is the number of color channels, (for RGB colorspace, $N_{\text{color}} = 3$), N_{frames} is the number of frames in the reconstructed frames and $N_{c,n}(l)$ is the number of subpixels originally set to l (but will be clipped to x_1) in the nth frame of color channel c in the reconstructed frames.

The ordered pair $(x_1, \mathrm{PSNR}(x_1))$ represents Operating Point 1 (OP_1) at which power may be saved by setting the BSF and SSF to x_1 and to $x_1^{-\gamma}$, respectively. The QoE loss at OP_1 is quantified by $\mathrm{PSNR}(x_1)$. As explained in Section 2.2, setting $S = x_{max}^{\frac{1}{\gamma}}$ maximizes power reduction with no quality loss. Therefore, the ordered pair $(x_{max}, \mathrm{PSNR}(x_{max}))$ is referred to as the No-Quality-Loss Operating Point (NQLOP). Note that $\mathrm{PSNR}(x_{max}) = \infty$.

2.3.2 A Power-Saving Protocol

During typical operation, the server determines $(N_Q + 1)$ OPs denoted by $OP_i = (x_i, \mathrm{PSNR}(x_i))$ for $i \in \{0, \ldots, N_Q\}$ and transmits them to the client as metadata. Note that OP_0 is the NQLOP. This enables the following power-saving protocol to be implemented at a mobile client. The user specifies a list of N acceptable PSNR quality levels $Q[1], \ldots, Q[N]$, where $Q[1] > Q[2] > \cdots > Q[N]$ and a list of Remaining Battery Life Levels (RBLLs) denoted $\mathrm{RBLL}[1], \ldots, \mathrm{RBLL}[N]$ so that $\mathrm{RBLL}[1] > \mathrm{RBLL}[2] > \cdots > \mathrm{RBLL}[N]$. For example, consider N = 3 and Q[1] = 40, Q[2] =

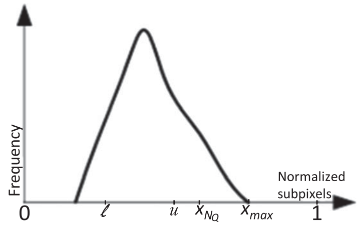

Fig. 4. Histogram showing lowest-quality OP (x_{N_Q}), NQLOP (x_{max}), lower bound (l), and upper bound (u).

35, Q[3] = 25 with RBLL[1] = 70 percent, RBLL[2] = 40 percent and RBLL[3] = 0 percent. When the user watches a video, the power optimization module monitors the actual RBLL, denoted $\text{RBLL}_{\text{actual}}$, of the device and selects RBLL [iSelected] so that $\text{RBLL[iSelected} - 1] > \text{RBLL}_{\text{actual}} > \text{RBLL[iSelected]}$, where RBLL[0] = 100 percent. For each frame to be displayed, the device examines the display-adaptation metadata and selects OP_j for which $Q[\text{iSelected} - 1] \geq \text{PSNR}[x_j] > Q[\text{iSelected}]$, where $Q[0] = \infty$. The metadata x_j is then used to determine the BSF, $x_j^{\frac{1}{\gamma}}$, and SSF, x_j^{-1}, that provide an appropriate QoE level.

2.3.3 Agnostic Metadata

One important requirement [16] for the Green Metadata standard is that the metadata should be agnostic about specific client implementations. In the preceding description on metadata generation and usage, observe that the server did not use any client-specific information (for example, the display gamma) to generate the metadata and therefore satisfied this requirement. However, at the client, where the display gamma is known, the power optimization module combines the metadata x_j with the client-specific display gamma γ to compute the client-specific BSF ($x_j^{\frac{1}{\gamma}}$) and SSF (x_j^{-1}) that provide power reduction. Thus, the client will implement a protocol that strikes the optimal balance between perceived quality and power-saving given the RBLL and user-specified PSNR levels. The balance is tilted toward quality when the RBLL is high but shifts toward power saving as the battery is depleted.

2.4 Metadata for Contrast Enhancement

2.4.1 Contrast Enhancement for Images

Cheng et al. [14] showed that aggressive backlight dimming can be done for *images* without excessive QoE loss if the contrast fidelity is maintained, rather than the perceived luminance. In this section, we explain how a modified version of their algorithm can provide Green Metadata for drastic power reduction. In Section 3, we show how the metadata should be computed for *video* sequences.

Given the $(N_Q + 1)$ OPs $(x_i, \text{PSNR}(x_i))$ for $i \in \{0, \ldots, N_Q\}$ for a reconstructed video frame, the standard provides contrast-enhancement metadata that enables a large power reduction at the lowest-quality, OP_{N_Q}. Consider the image histogram[2] in Fig. 4 wherein x_{N_Q} and x_{max} are respectively, the lowest-quality, OP_{N_Q}, and the NQLOP. At OP_{N_Q}, the BSF will be set to $S = x_{N_Q}^{\frac{1}{\gamma}}$ for power reduction. Typically, OP_{N_Q} will be selected for maximal power reduction so that $x_{N_Q} \ll x_{max}$. Consequently, many subpixels will be clipped after applying the SSF, $S^{-\gamma}$, and the QoE loss may be unacceptably high.

Fortunately, Cheng et al. [14] demonstrated that contrast enhancement can be used to improve the quality so that maximal power reduction is achievable with an acceptable QoE loss. The key idea is to reduce the subpixel dynamic range by truncating it on both ends, at lower-bound l and upper-bound u, in Fig. 4. An affine transformation is then applied to spread out the subpixels in the reduced range and thereby increase contrast. Let $C(x)$ perform the truncation and affine transformation as follows:

$$C(x) = \begin{cases} 0, & \forall x \in [0, l), \\ \frac{x-l}{u-l}, & \forall x \in [l, u], \\ 1, & \forall x \in (u, 1]. \end{cases} \tag{9}$$

The perceived luminance after the contrast enhancement is given by $\mathcal{L}_C = SB_M T(C(x))$

$$\mathcal{L}_C = \begin{cases} 0, & \forall x \in [0, l), \\ SB_M \left(\frac{x-l}{u-l}\right)^{\frac{1}{\gamma}}, & \forall x \in [l, u], \\ SB_M, & \forall x \in (u, 1]. \end{cases} \tag{10}$$

Comparing Equations (10) and (5), observe that the multiplicative SSF, $S^{-\gamma}$, which preserves luminance, has been replaced by the affine transformation $C(x)$. Also, from Equations (1) and (10), $\mathcal{L}_0 \neq \mathcal{L}_C$ and it is impossible to preserve luminance, for arbitrary l and u. But Cheng et al. have shown that the affine mapping $C(x)$ defines a contrast scaling that provides superior quality with larger power reduction than the multiplicative SSF, $S^{-\gamma}$, that preserves luminance. Their algorithm finds S, l and u that maximize power reduction for an *image* and cannot be used for *video* because S, l and u would change drastically between video frames causing flicker.

2.4.2 Contrast Enhancement Metadata for Video

For video, the server only provides contrast-enhancement metadata for the lowest quality, OP_{N_Q} and can thus assume that S is associated with this particular OP and therefore, S is fixed. To preserve luminance, the server would need to use the gamma-dependent BSF, $S = x_{N_Q}^{\frac{1}{\gamma}}$ defined in Section 2.3. However, as explained above, contrast scaling is superior to luminance preservation and consequently, the server uses the *gamma-independent*, fixed BSF $S = x_{N_Q}$ to determine l and u that maximize contrast thus providing acceptable QoE at the lowest quality OP_{N_Q} with maximum power reduction. Using the gamma-independent S at the server is consistent with the client-agnostic metadata requirement in Section 2.3.

In Equation (9), observe that the interval $[l, u]$ is mapped to the interval $[0, 1]$. Then, after applying the BSF, $S(= x_{N_Q})$ to the display, the interval $[l, u]$ is perceived visually as the interval $[0, SB_M]$. Therefore, for subpixels within the interval $[l, u]$, the Perceived Contrast Enhancement (PCE) is proportional to $SB_M/(u - l)$. Because B_M is a constant, we ignore it and quantify the PCE as PCE$= \frac{S}{u-l}$. For subpixels

2. Although Section 7 shows that Algorithm 1 performs well over a diverse set of video sequences, a more sophisticated approach may be required for images with multimodal, subpixel distributions.

within the intervals $[0, l)$ and $(u, 1]$, all contrast is lost because these intervals are mapped to 0 and 1, respectively, by Equation (9). From the preceding observations, it is clear that we should maximize the PCE by determining l and u to minimize $(u - l)$ while ensuring, simultaneously, that most subpixels lie within the interval $[l, u]$ to reduce loss of contrast. An additional constraint, that maintains contrast fidelity [14], is that PCE ≤ 1, so that $(u - l) \geq S$. Therefore, we use the following process to compute the contrast-enhancement metadata at the server. First, determine the BSF corresponding to the lowest quality level OP as $S = x_{N_Q}$. Then invoke the pseudocode function GetContrastMetadata() in Algorithm 1 to determine l and u.

Algorithm 1. $[l, u]$ = GetContrastMetadata(S,CDF, δ)

```
// S: BSF at the lowest quality OP_{N_Q}.
// CDF: cumulative distribution function for the subpixels.
// δ: step size for each search iteration.
maxEnhancement = 0
for (dr = S; dr < 1; dr = dr + δ) do
  for (l = 0; l < 1 − dr; l = l + δ) do
    u = l + dr;
    enhancement = (CDF(u) − CDF(l)) / (u − l)
    if (enhancement > maxEnhancement) then
      maxEnhancement = enhancement
      bestl = l;
      bestu = u;
    end if
  end for
end for
return (bestl, bestu)
```

After computing l and u for the lowest quality OP_{N_Q}, the server transmits all $(N_Q + 1)$ OPs $(x_i, \mathrm{PSNR}(x_i))$ for $i \in \{0, \ldots, N_Q\}$ along with the bounds l and u. When the client receives this metadata, its power optimization module will decide which OP to use, as explained in Section 2.3. If it chooses an OP_i for $i \in \{0, \ldots, N_Q - 1\}$, then the multiplicative SSF is applied as explained in Sections 2.2 and 2.3. However, at low RBLL, if the lowest quality OP indexed by $i = N_Q$ is selected, then the power optimization module maximizes power saving by using BSF, $S = x_{N_Q}$, along with Equation (9) that implements contrast scaling to provide acceptable quality with a huge power reduction. Without the contrast scaling, this huge power reduction would not be achievable at low RBLL.

3 Using Metadata for Video Sequences

3.1 What Causes Flicker in Video Display Adaptation?

In this section, we explain the additional processing and metadata that is required to reduce power consumption when displaying video in a server-client setting. The key challenge specific to video content is flicker reduction: at scene changes, dramatic variations in image characteristics cause dramatic metadata variations that result in flicker artifacts, which mar the viewing experience severely. For example, consider Fig. 5a in which the blue stems depict the BSFs over several frames before and after a scene change. This example shows a large increase in the BSF at the scene change. Conversely, the blue stems in Fig. 5b depict a large decrease in the BSF at the scene change. In either case, the large change in BSF at the scene change would cause flicker. Our solution is to prevent flicker by smoothing the BSFs temporally so that large variations are eliminated, as illustrated by the red stems in Fig. 5.

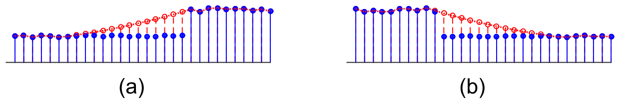

Fig. 5. Blue stems represent BSFs of successive frames. Red stems show temporally smoothed BSFs (a) after a large BSF increase; (b) after a large BSF decrease.

To explain the flicker artifacts that arise when metadata is naively applied to displays of video sequences, we introduce notation as follows. Consider a video sequence comprised of M frames, indexed by m, so that $m \in \{1, \ldots, M\}$. For each frame, the server determines BSFs that enable $(N_Q + 1)$ quality levels, indexed by i, so that $i \in \{0, \ldots, N_Q\}$. We denote the BSF for the mth frame at the ith quality level by $x_{i,m}$. For the ith quality level, let X_i represent the vector of M BSFs $[x_{i,1}, \ldots, x_{i,M}]$. For the mth frame, $(x_{i,m}, \mathrm{PSNR}(x_{i,m}))$ represents the ith operating point, where $i \in \{0, \ldots, N_Q\}$. Note that $\mathrm{PSNR}(x_{0,m}) > \mathrm{PSNR}(x_{1,m}) > \ldots > \mathrm{PSNR}(x_{N_Q,m})$ and $x_{0,m}$ is the largest subpixel in the mth frame. Therefore, $(x_{0,m}, \mathrm{PSNR}(x_{0,m}))$ is the NQLOP and $\mathrm{PSNR}(x_{0,m}) = \infty$. Let l_m and u_m represent, respectively, the lower- and upper-bound contrast-enhancement metadata associated with the lowest quality OP $(x_{N_Q,m}, \mathrm{PSNR}(x_{N_Q,m}))$ for the mth frame. Finally, vector L denotes $[l_1, \ldots, l_M]$ and vector U denotes $[u_1, \ldots, u_M]$.

3.2 Temporal Smoothing to Prevent Flicker

Now, under normal operating conditions in which the mobile client's battery level is fairly stable, the client will attempt to maintain a certain viewing QoE. At a high battery level, the client could choose to operate at the NQLOP with $i = 0$, for each frame, and thus experience the highest QoE with PSNR = ∞ for the BSFs X_0. When the battery level is low, the client could choose the lowest quality OP_{N_Q} and use BSFs X_{N_Q} to minimize battery drainage. The associated contrast-enhancement metadata L and U could also be used to improve quality at this OP. At intermediate battery levels, the client would determine i appropriately for $i \in \{1, \ldots, (N_Q - 1)\}$ and use the BSFs X_i.

Once i is fixed, the BSF $x_{i,m}$ will be applied to the mobile client's display for the mth frame. During homogeneous video scenes, the BSF will not change much from frame to frame. However, at scene changes, the video content and consequently, the BSF, may vary dramatically as shown in Fig. 5. Such large BSF variations may cause unpleasant flicker at scene changes. In our experiments, we determined that flicker is eliminated if the relative change between BSFs of successive frames is bounded. Specifically

$$\left| \frac{x_{i,m+1} - x_{i,m}}{x_{i,m}} \right| \leq \Delta_{BSF}, \tag{11}$$

for $i \in \{0, \ldots, N_Q\}$ and $m \in \{1, \ldots, (M - 1)\}$ with $\Delta_{BSF} \approx 0.015$. If the BSFs of successive frames, $x_{i,m}$ and $x_{i,m+1}$, do not satisfy the bound in Equation (11), then depending on their relative values, either $x_{i,m}$, or $x_{i,m+1}$, or both, may be decreased or increased until the bound is satisfied. To ensure that the NQLOP is maintained and that quality levels do not deteriorate, we devised Pseudocode 1 for temporal smoothing in which BSFs may be increased, but are never decreased. From Fig. 3c and Expression (8), observe that the perceived quality and PSNR both increase when the BSF, $x_{i,m}$ is increased. In particular, if $x_{0,m}$ is decreased to satisfy the bound in Equation (11), then $x_{0,m}$ will be smaller than the largest subpixel in the mth frame and X_0 will not provide the NQLOP for all frames, as required. Finally, note that if X_i satisfies the distortion constraint, then temporal smoothing by Pseudocode 1, will maintain this constraint because the algorithm never decreases BSFs.

Pseudocode 1. Smooth-BSFs(X_i, Δ_{BSF})

```
// X_i: vector of BSFs for M frames at the ith-quality OP
// Δ_BSF: bound on large relative variations in BSFs.
for (m = 2; m < (M + 1); m++) do
  Smooth-Temporally(X_i, Δ_BSF, m)
end for
```

Algorithm 2. Smooth-Temporally(MetaVec, Δ, m)

```
// MetaVec: vector with one metadata entry per frame
// Δ: bound on large relative variations in MetaVec
// m: indexes mth frame in MetaVec
if m == 1 then
  return
end if
cur = MetaVec[m]
prev = MetaVec[m − 1]
if |(cur−prev)/prev| > Δ then
  if cur < prev then
    // Increase current frame's metadata to meet bound.
    MetaVec[m] = prev × (1 - Δ)
  else
    // Increase previous frame's metadata to meet bound.
    // Then adjust metadata for all preceding frames.
    MetaVec[m − 1] = cur/(1+Δ)
    Smooth-Temporally(MetaVec, Δ, m − 1)
  end if
end if
```

The algorithm Smooth-BSFs prevents flicker by invoking the recursive algorithm Smooth-Temporally to adjust BSFs to meet the bound in Equation (11). However, at the lowest quality level with $i = N_Q$, flicker may also arise from large variations in the contrast-enhancement metadata, L and U. Fortunately, the recursive algorithm Smooth-Temporally can eliminate flicker from such large variations as shown in Pseudocode 2, which also invokes Smooth-Temporally. Note that Δ_{CE} is a display-independent bound on the relative variation in successive values of contrast-enhancement metadata such as l_m and l_{m+1}, or on u_m and u_{m+1}. Specifically

$$\left|\frac{l_{m+1} - l_m}{l_m}\right| \leq \Delta_{CE}, \tag{12}$$

$$\left|\frac{u_{m+1} - u_m}{u_m}\right| \leq \Delta_{CE}. \tag{13}$$

Our experiments have shown that no flicker occurs if the metadata in L and U satisfy the bounds in Equations (12) and (13) for $m \in \{1, \ldots, (M-1)\}$ with $\Delta_{CE} = 0.015$.

Pseudocode 2. Smooth-Contrast-Metadata(L, U, Δ_{CE})

```
// L = [l_1, ..., l_M]
// U = [u_1, ..., u_M]
// Δ_CE: bound defined in Equations (12) and (13).
for (m = 2; m < (M + 1); m++) do
  Smooth-Temporally(L, Δ_CE, m)
end for
for (m = 2; m < (M + 1); m++) do
  Smooth-Temporally(U, Δ_CE, m)
end for
```

4 Transmission of DA Metadata

The Green Metadata Standard specifies the syntax and the semantics for the DA metadata described in Sections 2 and 3. However, it does not provide any systems protocols for the transmission of this metadata from the server to the client. These protocols are being standardized in the two most widely used systems standards: the ISO Base Media File Format (ISO/IEC 23001-10) [2] and the MPEG-2 Transport Stream (ISO/IEC 13818-1). We shall now provide a high-level description for the transmission of Green Metadata via the ISO Base Media File Format (ISOBMFF). Transmission using the MPEG-2 Transport Stream follows a similar protocol to transmission via the ISOBMFF. We refer the interested reader to Reference [3] for details.

The ISOBMFF contains the timing, structure, and media information for timed sequences of media data, such as audio-visual presentations. In the ISOBMFF, the DA metadata is hierarchically structured and synchronized with corresponding video data. Transmission Format 1 shows the highest level. If $l > 0$, then the metadata provides contrast-enhancement bounds l and u, as defined in Section 2.4. The quality levels appear next in Transmission Format 2. As defined in Section 2.2, x_0 provides the BSF for the NQLOP with infinite PSNR. The loop then lists N_Q OPs denoted by OP_i for $i \in \{1, \ldots, N_Q\}$. Thus, the metadata provides the subpixel statistics, associated quality-level indicators and dynamic-range bounds that enable effective DA.

Transmission Format 1. Metadata-Set(N_Q)

```
N_Q // Number of Quality Levels. N_Q ∈ {1, ..., 15}
l // Lower contrast-enhancement bound
if l > 0 then
  u // Upper contrast-enhancement bound
end if
Quality-Levels(N_Q)
```

5 Complexity Analysis

To minimize power consumption on battery-powered mobile devices, the metadata extraction, subpixel scaling and power optimization module in Fig. 1 must use low-complexity algorithms. The metadata generator may use higher complexity algorithms because the server is not battery-powered and has powerful processors. We shall now discuss the complexity of the algorithms that operate on the server and on the client for metadata generation and application.

Transmission Format 2. Quality-Levels(N_Q)

x_0 // BSF that guarantees NQLOP with infinite PSNR
for $(i = 1; i \leq N_Q; i{+}{+})$ **do**
  x_i // BSF that provides the ith Quality OP
  PSNR(x_i) // PSNR of the ith Quality OP
end for

The metadata in Transmission Formats 1 and 2 are computed in the metadata generator, at the server. To determine x_0 in Transmission Format 2, the generator must compute the histogram in Fig. 4 and set $x_0 = x_{max}$. For a single frame, the histogram computation complexity is $O(WHN_{color})$, using the notation in Section 2.3. Typically, the N_Q quality levels in Transmission Format 2 are selected to allow graceful degradation from the highest quality level OP_0, to the lowest quality level, OP_{N_Q}. Accordingly, the N_Q quality levels are selected so that PSNR(x_i) is close to PSNR(x_{i+1}) for $0 \leq i < N_Q$. These N_Q quality levels may be determined from the histogram in Fig. 4 by selecting x_i and using Equation (8) to determine PSNR(x_i). Once the N_Q PSNRs are determined to ensure graceful degradation, the BSFs x_i may be obtained using a binary search of the histogram so that each OP has the desired PSNR. Therefore, computation of the N_Q quality levels has complexity $O(N_Q \log_2 P)$, where P is the largest subpixel value defined in Section 2.3. To obtain the contrast-enhancement bounds in Transmission Format 1, we use the nested loop in Algorithm 1 to determine the bounds from the CDF. Setting $K = \frac{1-S}{\delta}$, the complexity of Algorithm 1 is $O(K^2)$.

After obtaining the metadata for each frame, we use Pseudocodes 1 and 2 to smooth the BSFs and bounds temporally. Recall that each of these algorithms invokes the recursive Algorithm 2. For a sequence of M frames with D discontinuities of the type shown in Fig. 5, with largest relative variation V and with Δ bound on relative variations, the temporal smoothing complexity is $O(M + D\frac{V}{\Delta})$. If D is large, then the recursive call would be invoked frequently and the complexity would be high. Consequently, it is more efficient to generate metadata at a lower granularity by partitioning the video sequence into groups of frames. Transmission Formats 1 and 2 would then use histograms and CDFs computed over each frame group to generate metadata for each *frame group* instead of each *frame*, as described previously. The partitioning of frames into groups must ensure that scene changes occur at frame-group boundaries, and not within a frame group. Since the lighting conditions and pixel statistics are usually consistent within a scene, such a partitioning ensures that the metadata is appropriate for each frame in a group of frames. For a sequence of M frames partitioned into groups of G frames with d discontinuities of the type shown in Fig. 5, with largest relative variation v and with Δ bound on relative variations, the temporal smoothing complexity is $O(\frac{M}{G} + d\frac{v}{\Delta})$. Since $\frac{M}{G} \ll M$, $d < D$ and $v < V$, generating metadata at a lower granularity does reduce the complexity. Thus, the server experiences the following moderate computation load to process M frames partitioned into groups of G: $O(MWHN_{color} + \frac{M}{G} N_Q \log_2 P + \frac{M}{G} K^2 + \frac{M}{G} + d\frac{v}{\Delta})$

In contrast, the metadata extraction, subpixel scaling and power optimization module at the battery-powered mobile client have extremely low computational complexity. Specifically, metadata extraction requires the low-complexity parsing of a few bytes of data from an MPEG-2 TS or ISOBMFF bitstream. The subpixel scaling consumes negligible power because it merely scales each subpixel $x(i,j)$ to $x(i,j)S^{-\gamma}$ as shown in Equation (5). This operation dominates the client's computational load and has complexity $O(MWHN_{color})$. Finally, the power-optimization module determines the BSF and the SSF from metadata, display properties, pre-configured user settings and from the remaining battery life of the mobile client. This module consumes low power because it is infrequently invoked when the metadata changes at each scene or when the battery status falls below pre-determined thresholds.

6 Power-Adaptive Streaming

In addition to the metadata in Section 4, the ISOBMFF also provides related metadata that enables energy-efficient media selection [1], [4]. We now provide a brief explanation of how DA metadata contributes to this goal. To promote multimedia delivery over the internet, MPEG standardized Dynamic Adaptive Streaming over HTTP (DASH) [17] in 2011. This standard, and all preceding proprietary solutions, are based on client-side content selection based on network bandwidth monitoring: the client device selects the appropriate video and audio representations at each switching point as a function of the input bandwidth. The Green Metadata standard adds a new dimension to dynamic adaptive streaming: besides bandwidth considerations, media selection can be based on the media's power consumption impact on the client's decoder and display.

When the power resource is constrained, three media-selection strategies are possible:

1) Select a media representation that will require low power consumption. This leads to poor QoE because the lower resolution or lower bitrate representations will be selected.
2) For the next segment, select the most appropriate media representation based on the power consumption measured in the previous media segment. For example, on complex scenes (motion, detailed textures), the client device will select a representation with a lower resolution, while on simple scenes, it can switch back to a higher resolution. The issue is that the client device will react with at least one segment delay when the video-content complexity changes significantly. This leads to visible changes of perceived quality, particularly for simple scenes when the resolution is increased.
3) Use Green Metadata to select the most appropriate representation pro-actively by anticipating any change in the video content. This will guarantee the best QoE for a desired power-consumption level.

To enable the third strategy, the Green Metadata standard provides decoding-power and display-power indication metadata for each segment of a media stream. The display-power metadata is obtained by aggregating the DA metadata, as described in Section 4, over each segment.

Fig. 6. Experimental setup.

Therefore, this indication metadata describes the display quality levels for each segment, as well as the associated BSFs. When a client is ready to begin viewing a segment, it can use the following strategy to guarantee the best QoE for a desired power-consumption level.

First, the client determines the average acceptable power consumption based on its remaining battery life and the total duration of the video. Next, the client considers the resolutions, decoding-power consumptions and encoded quality levels of the available representations as well as the BSFs and display quality levels. Since the client can determine display power consumption from the BSFs, such aggregated DA metadata enables the client to select the most power-efficient representation with the best QoE.

7 Quality Assessment and Comparisons

7.1 MPEG Quality-Assessment Tests

The algorithms and metadata described in the preceding sections were submitted to the ISO/IEC MPEG. Over a two-year period, five quality assessment sessions were held to assess the performance of the technologies proposed to MPEG. At each viewing session, on average, 10 different MPEG experts examined live demos of the technologies. These viewers were industry experts with extensive prior training and experience in assessing video quality in consumer electronics. Following MPEG policy, if the consensus of the expert viewers was that the quality and power-reduction of a proposed technology was outstanding, then the technology would be adopted into the Green Metadata standard. Each of the five MPEG sessions were preceded by additional, rigorous, quality-assessment sessions conducted within the company that proposed the DA technologies. These sessions were each attended by approximately 20 engineers, executives and volunteers who critiqued the video quality. Thus, the technology has been evaluated scientifically by at least 50 different viewers over two years and more than 10 viewing sessions. In this section, we describe the quality-assessment test that was used for the adopted technologies in a video-streaming scenario.

7.1.1 Assessment Methodology

For the live demo in the quality test, we used the Samsung Galaxy Tab 2 platform (with a 1024x600 LCD panel) powered by the Monsoon Power Monitor (MPM) [18] which replaces the tablet's battery and logs instantaneous power consumption of the tablet, including power consumed by DA-processing. Our experimental setup is shown in Fig. 6. To ensure consistent results, we turned off all applications on the tablet and enabled airplane mode so that all radios were disabled. Under these conditions, the display will consume a larger fraction of overall power compared to normal operation conditions. Therefore, we expect that DA power reduction, under normal operating conditions, may be less than that reported in Section 7.1.2. To generate the metadata in Transmission Formats 1, 2, we used a C++ program running on a Samsung NP740U5M laptop with 7th Gen Intel Core i7-7500U processor and 16 GB RAM running Windows 10. On average, metadata generation consumed 2.5 msec/frame at 1080p resolution. Because our C++ program was not optimized, the reported metadata-generation time can be reduced significantly in an industrial deployment. We implemented SSF scaling in the tablet's Graphics Processing Unit's (GPU) OpenGL shader.

The test set consisted of 15 video sequences.[3] To determine the power reduction over the playback duration of each sequence, we connected the power monitor to the test platform so that it powers the platform and displays instantaneous power consumption. First, we played each sequence with the LCD backlight at its default (maximum) setting, L_0, and measured the average power-consumption, P_{ref}, over the playback duration. Next, as explained in Section 2.3, we set the BSF and the SSF for the highest-quality OP $(x_1, \mathrm{PSNR}(x_1))$, recorded the Structural Similarity (SSIM) quality measurement, played the sequence on the platform and took an average power-consumption reading, P_{testHi}, over the playback duration.[4] The power-reduction percentage at this OP is $(100 \times \frac{P_{ref}-P_{testHi}}{P_{ref}})$. In this methodology, the highest-quality OP is selected such that $\mathrm{PSNR}(x_1) \approx 40$ dB so that moderate power reduction is obtained at high QoE: this OP is *not* the NQLOP. Finally, we set the BSF and the SSF for the lowest-quality OP $(x_{N_Q}, \mathrm{PSNR}(x_{N_Q}))$, and re-played the sequence with contrast enhancement based on the l and u bounds in Section 2.4. If P_{testLo} denotes the average power-consumption reading over this playback duration, then $(100 \times \frac{P_{ref}-P_{testLo}}{P_{ref}})$ is the power-reduction percentage at the lowest-quality OP.

7.1.2 Assessment Results: MOS, PSNR, and SSIM

The SSF scaling in the tablet's GPU incurred negligible power/latency overhead because we did not observe any such overheads in our experimental setup even though 10 of the test sequences ran at 30 fps or higher. In Table 1, the PSNR and SSIM are objective, industry-standard, quality-assessment metrics that enable comparison between power-reduction from the proposed method and from Lin et al.'s method [11]. The left half of the table shows the power reduction at the highest- and lowest-quality OPs. Quality metrics are not shown for the lowest-quality OP because they are unreliable after applying contrast enhancement. Instead, we relied on visual inspection (e.g., Fig. 7) to assess the lowest-quality OP. For all sequences, the viewers examined the videos carefully and confirmed that the highest-quality OP had

3. Eleven of these sequences are described in the MPEG HEVC CfP [19] because they were used for the development of that standard. The respective spatial resolution, frame rate (in frames per second) and duration (in seconds) for the remaining four sequences are as follows: Thor (1920x816, 23.976, 83), WhiteCollar (1280x720, 23.976, 62), Aurora (1280x720, 30, 115), and OldTownCross(1920x1080, 25, 20). All sequences were compressed using MPEG AVC.

4. The frame groups in each sequence were determined manually so that scene changes occurred at frame-group boundaries. Section 8 explains how this can be done automatically.

TABLE 1
Power Reduction Measurements: PR = Power Reduction, LQ = Lowest Quality, PSNR = Peak Signal to Noise Ratio, and SSIM = Structural SIMilarity

		PR Measurements (Proposed Method)				PR Measurements for [11]				Minimum
		Highest Quality			LQ	Highest Quality			LQ	Backlight
Sequence	P_{ref} (mW)	PR %	PSNR (dB)	SSIM	PR %	PR %	PSNR (dB)	SSIM	PR %	PSNR (dB)
Vidyo1	3052	17.3	42.1	0.9945	75.1	17.3	29.2	0.9952	75.7	10.0
BasketballDrill	3047	28.2	41.3	0.998	91.7	30.1	24.0	0.9794	93.1	10.6
BasketballPass	2951	35.4	43.1	0.9993	90.6	36.0	21.4	0.9629	91.8	10.5
BlowingBubbles	2952	18.0	41.6	0.9967	91.2	21.2	27.7	0.9917	92.0	10.6
BQMall	3021	21.5	40.0	0.9981	63.2	22.6	25.8	0.9869	64.6	10.5
Kimono	3230	27.8	42.3	0.995	88.8	30.6	25.4	0.9691	90.0	14.8
PartyScene	3032	18.5	41.7	0.999	82.7	18.5	28.6	0.981	84.2	10.2
RaceHorses	2905	19.8	41.4	0.9956	95.3	25.5	27.1	0.9915	96.5	10.4
RaceHorsesC	2987	18.6	43.5	0.9955	93.0	21.5	28.8	0.9917	94.1	10.3
Vidyo3	3073	68.4	40.7	0.9964	90.1	68.4	16.3	0.9965	91.7	9.9
BQSquare	2950	9.6	41.7	0.9906	51.9	9.6	31.2	0.9980	53.2	6.7
Aurora	3045	15.7	38.9	0.9987	74.7	20.5	26.5	0.9709	95.1	12.3
OldTownCross	3363	43.8	42.8	0.9993	89.9	43.8	18.7	0.9554	88.5	9.6
Movie Clip (Thor)	3148	93.6	42.8	0.9918	94.0	97.7	22.8	0.5071	77.6	23.1
TV Clip (WhiteCollar)	3047	51.4	41.0	0.9866	88.2	58.7	17.4	0.8071	91.2	19.5
Average	**3054**	**32.5**	**41.7**	**0.9957**	**84.0**	**34.8**	**24.7**	**0.9390**	**85.3**	**11.9**
Std. Deviation	**118**	**23.1**	**1.2**	**0.0036**	**12.6**	**23.6**	**4.5**	**0.1284**	**12.4**	**4.2**

excellent quality (Mean Opinion Score (MOS) = 5, on 5-point scale[5]) and that the lowest-quality OP had moderate quality (MOS = 3) which would be acceptable when the mobile device's battery charge was low. The viewers also confirmed that for each sequence, unacceptable quality (MOS = 2) was observed if the OP was set below the lowest-quality OP listed for that sequence in Table 1. In particular, the quality was completely unacceptable for all sequences when the backlight was set to its minimum setting of 10 percent (MOS = 1). The average PSNR for this setting was 11.9 dB as shown in the column labeled "Minimum Backlight" in Table 1. Pasricha et al. [9] made the same observation for their "Simple Backlight Compensation" system which operates similarly.

Observe that the Thor movie clip provides much greater power reduction than the other 14 sequences. This happens because the movie clip contains bright as well as dark scenes. Consequently, the dark scenes trigger large power reductions since the BSF, and hence, the power consumption, is low during such scenes. In contrast, all the other sequences (besides Aurora and WhiteCollar) were selected to test video compression tools [19] and they are uniformly bright with high contrasts. In particular, BQSquare has the lowest power reduction because a large percentage of its pixels are bright. Note that all reported power reductions are relative to the power consumption of *the entire device* and are not limited to the display unit within the device. Therefore, any power consumption from DA-processing on the client *is accounted for in the measurements*. Throughout the experiment, we did not observe any unexpected negative effects from DA-processing.

5. These quality assessment results are from the final test conducted by MPEG. At this stage, the DA technology had been improved based on feedback from earlier tests. Consequently, all viewers concurred on the quality levels and reported identical MOS scores. Therefore, because they are identical, MOS scores are not reported separately in Table 1. We used the following MOS scale: 5 = excellent (imperceptible impairment), 4 = good (perceptible but not annoying impairment), 3 = fair (slightly annoying impairment), 2 = poor (annoying impairment), and 1 = bad (very annoying impairment).

7.1.3 Assessment Results: Contrast-Enhancement

When the dynamic-range bounds l and u are used for contrast enhancement at the lowest-quality OP, we measure up to 84 percent power reduction at acceptable quality levels (MOS = 4). Fig. 7 illustrates the improvement from these bounds. Note that this figure simulates the backlit LCD display and accurately represents the viewing experience during the previously described quality assessment. Specifically, Figs. 7a, 7b and 7c simulate the backlit LCD display using Equation (5) while Fig. 7e uses Equation (10). We encourage the reader to view Fig. 7 on an LCD monitor rather than on printed paper. Fig. 7a shows a video frame from the OldTownCross sequence, displayed under the default full backlight, B_M. Fig. 7c shows the frame at the highest-quality OP with 43.8 percent power reduction and virtually no QoE loss. Fig. 7b shows the frame at the lowest-quality OP with 89.9 percent power reduction, without contrast enhancement. It appears washed out because of heavy pixel saturation at a low value of x_N. During the viewing test, the consensus was that Fig. 7e depicts a quality improvement from using the l and u bounds for contrast enhancement at the lowest-quality OP. The power reduction after contrast enhancement is still 89.9 percent.The viewers noted that many objects were invisible in Fig. 7b, but clearly visible in Fig. 7e. Specifically, in Fig. 7b, the clouds, white cars on the road and buildings silhouetted against the sky are invisible. However, these objects are all clearly visible in Fig. 7e. Although Fig. 7e appears dark, the consensus was that it was preferable to the washed out Fig. 7b which hurt the eyes on a mobile LCD display. The viewers made similar observations for other video sequences besides OldTownCross.

Fig. 7. (a) OldTownCross frame under full backlight, B_M. (b) Lowest-quality OP without contrast enhancement and 89.9 percent power reduction. (c) Highest-quality OP with 42.8 percent power reduction. (d) Highest-quality OP using [11] and 43.8 percent power reduction. (e) Lowest-quality OP with contrast enhancement and 89.9 percent power reduction. (f) Lowest-quality OP using [11] and 91.2 percent power reduction.

7.1.4 Comparison with Lin et al. [11]

As explained in Section 1, Lin et al. [11] deployed a server-client DA service based on their algorithm that computes optimal BSFs and deploys them on mobile devices without SSF scaling and without contrast enhancement. In Columns 7, 8, 9, 10 of Table 1, we provide power-reduction, PSNR and SSIM measurements for their approach, using the assessment methodology in Section 7.1.1. Viewing confirmed that Lin's method provides good QoE with slightly higher power reduction compared to the high quality OP from our proposed method in Columns 3, 4, 5 of Table 1. Since Lin's method does not use SSF scaling, the reported PSNRs are low, although the QoE is acceptable. Lin's method has low SSIM on dark sequences (e.g., Thor) because, without SSF scaling, many subpixels are close to zero and reduce the SSIM. Fig. 7d shows the video frame from the OldTownCross sequence, displayed at the highest-quality OP using a backlight computed using Lin's method at 43.8 percent power reduction and slight QoE loss from the lack of SSF scaling. However, at the lowest-quality OP with large power reductions, Lin's method has a low QoE as seen in Fig. 7f with 91.2 percent power reduction. This happens because their method does not utilize contrast-enhancement metadata. As explained in Section 8, Lin's method can be used instead of Algorithm 1 to generate standardized metadata because the Green Metadata standard does not standardize algorithms.

7.1.5 Assessment Results: Flicker Analysis

At the viewing test, the viewers also noted that video playback occurred without any flicker at all OPs when the metadata was generated with $\Delta_{BSF} < 0.015$ and $\Delta_{CE} < 0.015$.

TABLE 2
Flicker Assessment Comparison

Sequence	Flicker Assessment Comparison				
	Proposed Method			Raman et al. [15]	
	Flicker with C1	Flicker with C2	Distortion Constraint	Flicker with smoothing	Distortion Constraint
Vidyo1	No	No	Satisfied	No	Violated
BasketballDrill	No	No	Satisfied	No	Violated
BasketballPass	No	No	Satisfied	No	Violated
BlowingBubbles	No	No	Satisfied	No	Violated
BQMall	No	No	Satisfied	No	Violated
Kimono	No	No	Satisfied	No	Violated
PartyScene	No	No	Satisfied	No	Violated
RaceHorses	No	No	Satisfied	No	Violated
RaceHorsesC	No	No	Satisfied	No	Violated
Vidyo3	No	No	Satisfied	No	Violated
BQSquare	No	No	Satisfied	No	Violated
Aurora	Yes	No	Satisfied	No	Violated
OldTownCross	No	No	Satisfied	No	Violated
Movie Clip (Thor)	Yes	No	Satisfied	No	Violated
TV Clip (WhiteCollar)	Yes	No	Satisfied	No	Violated

Under Condition 1 (C1), metadata is generated with $\Delta_{BSF} > 0.015$ *and* $\Delta_{CE} > 0.015$. *Under Condition 2 (C2), metadata is generated with* $\Delta_{BSF} < 0.015$ *and* $\Delta_{CE} < 0.015$

These observations are logged in the right half of Table 2 in the column headed "Flicker with Condition 2". However, when metadata was generated with these thresholds set larger than 0.015, then flicker did occur at all OPs in Aurora, Thor and White Collar sequences, as logged in the column headed "Flicker with Condition 1". These sequences contained several heterogeneous scenes and therefore the large thresholds enabled large, inter-frame BSF changes that caused flicker. In the normal operation mode, as explained by Equations (11), (12), (13), metadata will be generated under Condition 2, and flicker is absent. Since the BSFs are smoothed temporally by Pseudocode 1 which maintains the distortion constraint, it is impossible to obtain a low-quality frame after temporal smoothing. This fact is recorded in the column headed "Distortion Constraint" for the proposed method.

7.1.6 Comparison with Raman et al. [15]

As explained in Section 1.1, Raman et al. [15] provide the only state-of-the-art technique for display adaptation with *video* contrast enhancement. We operated their algorithm at the same OPs as our proposed method and measured the same power reductions, since the BSFs were identical and not changed much by their smoothing algorithm. Viewing confirmed that no flicker was observed on any sequence, as logged under "Flicker with smoothing" in Table 2. However, because their algorithm uses exponential smoothing, BSFs are decreased, thus violating the distortion constraint as recorded in the column headed "Distortion Constraint" for Raman et al.

7.2 Comparison against Related Techniques

The ecosystem that supports video consumption on mobile devices consists generally of entities that create, transport and display video. From each entity's perspective, we now evaluate the Green Metadata standard objectively and compare it against all related Display Power-Reduction (DPR) techniques which are categorized into Non Server-Client (NSC) methods [13], [14] and Server-Client (SC) methods [9], [10], [11], [12].

In Table 3, we compare NSC methods, SC methods and Green Metadata from the perspective of three entities in ecosystems that enable video streaming (e.g., YouTube, Netflix), sharing (e.g., Google+, Facebook), live streaming (e.g., YouTube, Facebook), conferencing (e.g., Skype, FaceTime). These three entities perform video creation, transport and display. The creation entity generates video content and generally resides on a server. The transport entity is usually a network that carries the video content from the creation entity to the mobile device. The display entity resides on a mobile device and renders the video content that was delivered by the transport entity. Video sharing is conceptually identical to streaming. The only difference is that in the sharing application, the video content originates on a mobile device and is uploaded to a server. After this upload, the content is prepared for streaming and the subsequent analysis is identical to the analysis of video streaming. Therefore we analyze these two applications together in the table. Similarly, live streaming and conferencing are analyzed together because they both use real-time video feeds. Note that conferencing is more demanding than live streaming because it has a lower latency requirement.

In NSC methods, all DPR-related processing is performed by the display entity on the mobile device. Therefore, the display entity incurs high computational complexity related to DPR, the creation entity incurs no DPR-related computational load and the transport entity is unaffected by DPR. In SC approaches, for the streaming application, the creation

TABLE 3
Tabular Comparison of Non-Server Client Methods, Server-Client Methods, and Green Metadata

Application	Entity	Non Server-Client Methods	Server-Client Methods	Green Metadata
Streaming/Sharing	Creation	Zero complexity	High complexity Closed ecosystem	High complexity Open ecosystem
	Transport	N/A	No synchronization No power-adaptive streaming	Synchronization Power-adaptive streaming
	Display	High complexity	Low complexity CE violates DC	Low complexity CE satisfies DC
Conferencing/Live-Streaming	Creation	Zero complexity	Moderate complexity Closed ecosystem	Moderate complexity Open ecosystem
	Transport	N/A	No synchronization	Synchronization
	Display	High complexity	Low complexity CE violates DC	Low complexity CE satisfies DC

Acronyms: CE = Contrast Enhancement, DC = Distortion Constraint, N/A = Not Applicable.

entity performs most of the DPR-related processing and thus has high computational complexity while the display entity enjoys low complexity from DPR-related processing. The creation and display entities must co-exist in a closed ecosystem, which is not true for most ecosystems that support streaming and sharing. The Green Metadata standard enables a SC approach that uses standardized metadata in open ecosystems with high complexity at the creation entity and low complexity at the display entity (Section 5). Furthermore, since the metadata is stored and carried using standardized transport protocols (MPEG-2, ISOBMFF), the transport entity can synchronize the metadata with the video content (Section 4). The transport entity can also enable power-adaptive streaming (Section 6). From the perspective of the display entity, Green Metadata provides the only approach for video contrast enhancement without flicker while maintaining the distortion constraint (Section 7).

For the video conferencing application, the creation and display entities may both reside in distinct, mobile devices. Because of stringent latency requirements and to avoid excessive battery depletion, the creation entity must perform moderate-complexity DPR-related processing, as explained in Section 8. Therefore, the creation entity incurs moderate computational complexity for the conferencing application in SC and Green Metadata approaches. For the live-streaming application, although the latency requirement is not as low as that of the conferencing application, live-streaming should still have lower latency than the streaming and sharing applications. So the live-streaming creation entity also incurs moderate computational complexity.

8 Implementation Issues

During the standardization process, the algorithms that we have described in this paper were selected as simple, pedagogical examples because they could be explained easily in the standard. For example, although Algorithm 2 is suboptimal, it was simple to explain and pedagogically valuable because it could be used to smooth both BSFs and contrast metadata. In an industrial deployment, other optimal algorithms (e.g., [11], [12]) may be used for metadata generation/usage. Specifically, in Fig. 1, only the format of the Green Metadata has been standardized. However, the algorithms in the Green Metadata generator and in the power optimization module, are *not* standardized: servers may use any algorithms that output Green Metadata in the standardized Transmission Formats 1, 2.

To address the implementation issues in an open-ecosystem DA deployment, consider Fig. 1. As explained in Section 5, the Green Metadata Generator (GMG) must output metadata for frame groups. The GMG may use a scene-change detector to determine the start and end of each group. Thus the group size is content dependent. The GMG could augment the scene-change detector with a video-scene segmentation algorithm that ends a frame group when the statistics of the next frame differ significantly from those of the preceding group.

For the video-conferencing application scenario described in Section 7.2, the GMG must perform moderate-complexity processing because it operates in a battery-powered mobile device with a low-latency requirement. From Section 5, recall that the server experiences the following complexity to process M frames partitioned into groups of G: $O(MWHN_{color} + \frac{M}{G} N_Q \log_2 P + \frac{M}{G} K^2 + \frac{M}{G} + d \frac{v}{\Delta})$. Now, the low-latency requirement restricts M to small values and thus reduces GMG processing complexity. Decreasing M also indirectly reduces the complexity because lowering M also diminishes d since smaller blocks of frames will have fewer discontinuities of the type shown in Fig. 5a. To reduce the complexity further, the GMG may use large values of G (satisfying $G \leq M$) and small values of N_Q. Although the low-latency requirement forces the GMG to compute metadata more frequently (due to small values of M), the metadata bitrate does not increase because the GMG needs to transmit metadata only when there is a significant change from previously transmitted metadata.

To estimate the bandwidth needed for the metadata, note that, in Section 4, N_Q, l, u, x_0, x_i, $\mathrm{PSNR}(x_i)$ require 8 bits each and, typically, $N_Q = 5$. So, the total number of bits in a metadata set is $8 + 8 + 8 + 8 + 5(8 + 8) = 32 + 5 \times 16 = 32 + 80 = 112$ bits. Assuming that each frame group lasts for 4 seconds, on average, and has an associated 112-bit metadata set, the typical metadata bandwidth is $112/4 = 28$ bps. To transport such metadata, the GMG encapsulates it in ISOBMFF or MPEG-2 thus synchronizing the metadata with the associated video bitstream, as mentioned in Table 3. Consequently, network quality issues do not affect the metadata: if the network maintains enough bandwidth for the video bitstream then the ISOBMFF / MPEG-2 synchronization ensures that the much smaller metadata will be delivered reliably to the client.

If the client uses power-adaptive streaming, then an interrupt is raised when the battery level reaches predetermined states (e.g., 60, 30, 10 percent). An Interrupt Service Routine (ISR) examines the decoding-power and display-power indication metadata and selects the best representation for the new battery state. A detailed discussion of the decoding-power indication metadata is beyond the scope of this paper. A lookup table maps the decoder's processor voltage/frequency to power consumption and allows the client to select the appropriate decoder voltage/frequency. Further details are in References [1], [4]. Because the ISR runs infrequently, it incurs negligible power/latency overheads and consequently, power-adaptive streaming selects the best representation for each battery-level state.

On the client, SSF scaling should be performed in the GPU because this incurs negligible power/latency overheads, as observed in Section 7.1.2. For LCD's, the Green Metadata enables large power-reductions because power consumption is proportional to the BSF. However, for OLED's, power consumption is also dependent on the SSF and, consequently, OLED power-reduction is a research topic [20], [21] that is currently under investigation for standardization in a future version of the Green Metadata standard. As such, metadata-based OLED DA is beyond the scope of this paper which describes the current version of the standard [1].

9 Conclusion

The Green Metadata in Fig. 1 consists of subpixel statistics, associated quality indicators and dynamic-range bounds that enable DA with an optimal power versus quality trade-off, and without flicker on video content. Because the DA metadata is standardized, the generation, storage, carriage, extraction and application of the metadata can be performed by independent entities in open ecosystems. We have demonstrated significant power reductions of 32.5 percent at high quality, and up to 84 percent at acceptable quality on LCD. The French GreenVideo project [22] is currently considering deployment of the Green Metadata standard.

Acknowledgments

F. C. Fernandes, Z. Ma, and X. Li contributed to this research when they were at Samsung Research America. Dr. Z. Ma is supported in part by the National Natural Science Foundation of China (Grant # 61571215).

References

[1] "Information Technology - MPEG Systems Technologies - Part 11: Energy-Efficient Media Consumption (Green Metadata)," *ISO/IEC JTC1/SC29/WG11 23001–11*, Oct. 2014.

[2] "Carriage of Timed Metadata Metrics of Media in the ISO Base Media File Format," *ISO/IEC JTC1/SC29/WG11 23001–10*, Oct. 2014.

[3] "Study Text of DAM 3 to ISO/IEC 138181:2014 - Carriage of Green Metadata in MPEG-2 Systems," *ISO/IEC JTC1/SC29/WG11 13818–1:2014*, Geneva, Switzerland, Feb. 2015.

[4] F. C. Fernandes, X. Ducloux, Z. Ma, E. Faramarzi, P. Gendron, and J. Wen, "The green metadata standard for energy-efficient video consumption," *IEEE MultiMedia*, vol. 22, no. 1, pp. 80–87, Jan.–Mar. 2015.

[5] N. Chang, I. Choi, and H. Shim, "DLS: Dynamic backlight luminance scaling of liquid crystal display," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 12, no. 8, pp. 837–846, Aug. 2004.

[6] A. Iranli, H. Fatemi, and M. Pedram, "HEBS: Histogram equalization for backlight scaling," in *Proc. Des. Autom. Test Europe*, 2005, pp. 346–351.

[7] A. Iranli, W. Lee, and M. Pedram, "Backlight dimming in power-aware mobile displays," in *Proc. 43rd Annu. Des. Autom. Conf.*, 2006, pp. 604–607.

[8] D. Shin, Y. Kim, N. Chang, and M. Pedram, "Dynamic voltage scaling of OLED displays," in *Proc. 48th ACM/EDAC/IEEE Des. Autom. Conf.*, 2011, pp. 53–58.

[9] S. Pasricha, M. Luthra, S. Mohapatra, N. Dutt, and N. Venkatasubramanian, "Dynamic backlight adaptation for low-power handheld devices," *IEEE Des. Test Comput.*, vol. 21, no. 5, pp. 398–405, Sep./Oct. 2004.

[10] L. Cheng, S. Bossi, S. Mohapatra, M. E. Zarki, N. Venkatasubramanian, and N. Dutt, "Quality adapted backlight scaling (QABS) for video streaming to mobile handheld devices," in *Proc. Int. Conf. Netw.*, 2005, pp. 662–671.

[11] C.-H. Lin, P.-C. Hsiu, and C.-K. Hsieh, "Dynamic backlight scaling optimization: A cloud-based energy-saving service for mobile streaming applications," *IEEE Trans. Comput.*, vol. 63, no. 2, pp. 335–348, Feb. 2014.

[12] P.-C. Hsiu, C.-H. Lin, and C.-K. Hsieh, "Dynamic backlight scaling optimization for mobile streaming applications," in *Proc. 17th IEEE/ACM Int. Symp. Low-Power Electron. Des.*, 2011, pp. 309–314.

[13] T.-H. Huang, K.-T. Shih, S.-L. Yeh, and H. Chen, "Enhancement of backlight-scaled images," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4587–4597, Dec. 2013.

[14] W.-C. Cheng and M. Pedram, "Power minimization in a backlit TFT-LCD display by concurrent brightness and contrast scaling," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 25–32, Feb. 2004.

[15] N. Raman and G. J. Hekstra, "Content based contrast enhancement for liquid crystal displays with backlight modulation," *IEEE Trans. Consum. Electron.*, vol. 51, no. 1, pp. 18–21, Feb. 2005.

[16] "Context, objectives, use cases and requirements for Green MPEG," MPEG Document N13468, *ISO/IEC JTC1/SC29/WG11*, Incheon, Korea, Apr. 2013.

[17] "Information technology Dynamic adaptive streaming over HTTP (DASH) Part 1: Media presentation description and segment formats," *ISO/IEC JTC1/SC29/WGLL*, 2012.

[18] Monsoon power monitor. (2018). [Online]. Available: http://www.msoon.com/LabEquipment/PowerMonitor/

[19] "Joint call for proposals on video compression technology," MPEG Document N11113, *ISO/IEC JTC1/SC29/WG11*, Kyoto, Japan, Jan. 2010.

[20] D. Shin, Y. Kim, N. Chang, and M. Pedram, "Dynamic driver supply voltage scaling for organic light emitting diode displays," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 32, no. 7, pp. 1017–1030, Jul. 2013.

[21] M. M. Dai and V. Lachine, "Perception optimized signal scaling for OLED power saving," in *Proc. SID Symp. Dig. Tech. Papers*, 2014, pp. 732–735. [Online]. Available: http://dx.doi.org/10.1002/j.2168–0159.2014.tb00192.x

[22] The French GreenVideo Project. (2018). [Online]. Available: http://greenvideo.insa-rennes.fr/

Felix C. Fernandes received the PhD degree in electrical engineering from Rice University. From 2010 to 2016, he was director of media-processing research and standardization with Samsung Research America, Dallas. His team developed and standardized new technology for video coding and fingerprinting, image search, multimedia transport, and energy-efficient processing. During this period, he chaired the ISO/IEC Green MPEG AdHoc Group and served as editor of the MPEG Green Metadata Standard. Currently, he is a video-codec architect with Apple.

Esmaeil Farmarzi received the BS and MS degrees from the Amirkabir University of Technology, Tehran, Iran, in 2000 and 2003, respectively, and the PhD degree from Southern Methodist University, Dallas, Texas, in 2012. From 2003 to 2008, he was a research faculty member with the Iranian Research Institute for Information Science and Technology (IRANDOC), Tehran, Iran. Since 2012, he has been with Samsung Research America, Richardson, Texas. His research experiences are in the areas of virtual reality, super resolution, texture synthesis/inpainting, display adaptation, document image analysis, optical character recognition, and image indexing and retrieval.

Xin Li received the BS and MS degrees from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003 and 2005, respectively, and the MS degree from the Volgenau School of Engineering, George Mason University, Fairfax, Virginia, in 2013. He is now the CTO of a startup company with focuses on cloud gaming and gigapixel streaming. From 2012 to 2015, he worked with Samsung Research America on Green MPEG, MicroServer, and the Samsung Virtual Reality platform. From 2005 to 2008, he worked with the Intel China Software Development Center at Shanghai on UEFI - the next generation BIOS.

Zhan Ma (S'06-M'11) received the BS and MS degrees from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2004 and 2006, respectively, and the PhD degree from the Tandon School of Engineering, New York University (formerly Polytechnic University, Brooklyn, New York), New York, in 2011. He is now on the faculty of the Electronic Science and Engineering School, Nanjing University, Jiangsu, 210093, China. From 2011 to 2014, he was with Samsung Research America, Dallas, Texas, and Futurewei Technologies, Inc., Santa Clara, California, respectively. His current research focuses on the multispectral video processing, Gigapixel streaming, and FPGA implementation. He is a member of the IEEE.

Xavier Ducloux received the graduate degree in telecommunications from the National Institute of Applied Science, France, in 1990. He is a research program manager with the Innovation group of Harmonic. With 22 years of experience in the video compression and digital TV industry, he is in charge of advanced studies through research partnerships and is active in MPEG standardization, where he is currently chairing Green MPEG AhG. He started his career working in radio propagation simulations for the French Defense Agency (DGA). He joined the Thomson group, in 1994, where he led video compression algorithms studies for several generations of HDTV encoder products and was involved in several European projects. He holds many patents in video compression and is particularly focused on the topic of video quality. He graduated from the National Institute of Applied Science in France in 1990 with a degree in telecommunications.