The Green Metadata Standard for Energy-Efficient Video Consumption

Felix C. Fernandes
Samsung Research America

Xavier Ducloux
Thomson Video Networks, France

Zhan Ma
Huawei Technologies

Esmaeil Faramarzi
Samsung Research America

Patrick Gendron
Thomson Video Networks, France

Jiangtao Wen
Tsinghua University, China

If you use your mobile device for video applications, then you probably wish that your mobile battery charge lasted longer. Given the current popularity of video applications such as streaming (YouTube, Netflix), conferencing (Skype, Google Hangout) and sharing (Google+, Facebook), there is already a huge demand for increased battery life on mobile devices. This demand will only intensify because of the anticipated 69.1 percent annual growth rate of mobile video traffic from 2013 to 2018.[1] Unfortunately, innovations in battery research are unlikely to satisfy the demand for longer battery life during mobile video consumption. The only solution is to provide energy-efficient video consumption.

In April 2013, MPEG issued a call for proposals (CFP) on energy-efficient video consumption. The CFP responses showed that, without any loss in the quality of experience (QoE), metadata can be used to reduce power consumed during video encoding, decoding, display, and selection. Furthermore, when battery levels are critically low, consumers may use the metadata to trade-off between QoE and larger power savings.

Based on compelling evidence from the CFP responses, MPEG initiated standardization of Green Metadata[2] for energy-efficient video consumption. This article describes how metadata enables large power reductions when QoE is maintained and even larger reductions when QoE is allowed to vary. When QoE is maintained, metadata enables average power reductions of 12, 12, and 26 percent during encoding, decoding, and display, respectively. In addition, we measured up to 80 percent power savings at lowered, but acceptable, QoE levels. In this article, we describe the functional architecture of a system that exploits the Green Metadata standard for energy-efficient media consumption, referring to this functional architecture to explain power reductions in various system components.

Editor's Note

Video consumption on mobile devices is increasing at a rapid pace, and there is a greater need to realize techniques that could save power. This article provides an overview of a newly developed standard for Green Metadata that enables energy-efficient media consumption.

A Functional Architecture for Green Metadata

Figure 1 shows the functional architecture for a transmitter/receiver system that uses Green Metadata, which is generated during preprocessing and/or encoding at the transmitter. The metadata is then multiplexed with or encapsulated into a bitstream that is delivered to the receiver. The metadata extractor sends the Green Metadata to a power-optimization module that interprets it and then applies controls to reduce power consumption during decoding and display of the video.

In addition, the receiver's power-optimization module can collect receiver information, such as remaining battery capacity, and send it to the transmitter as *green feedback* that adapts the encoder operations for power-consumption reduction.

Display Power Reduction

Display units in modern electronic devices consume large amounts of power. Therefore, display power reduction is a critical problem, especially for mobile devices with limited battery charges. The *display adaptation* (DA) technique (also known as *backlight dimming*)[3,4] has

1070-986X/15/$31.00 © 2015 IEEE

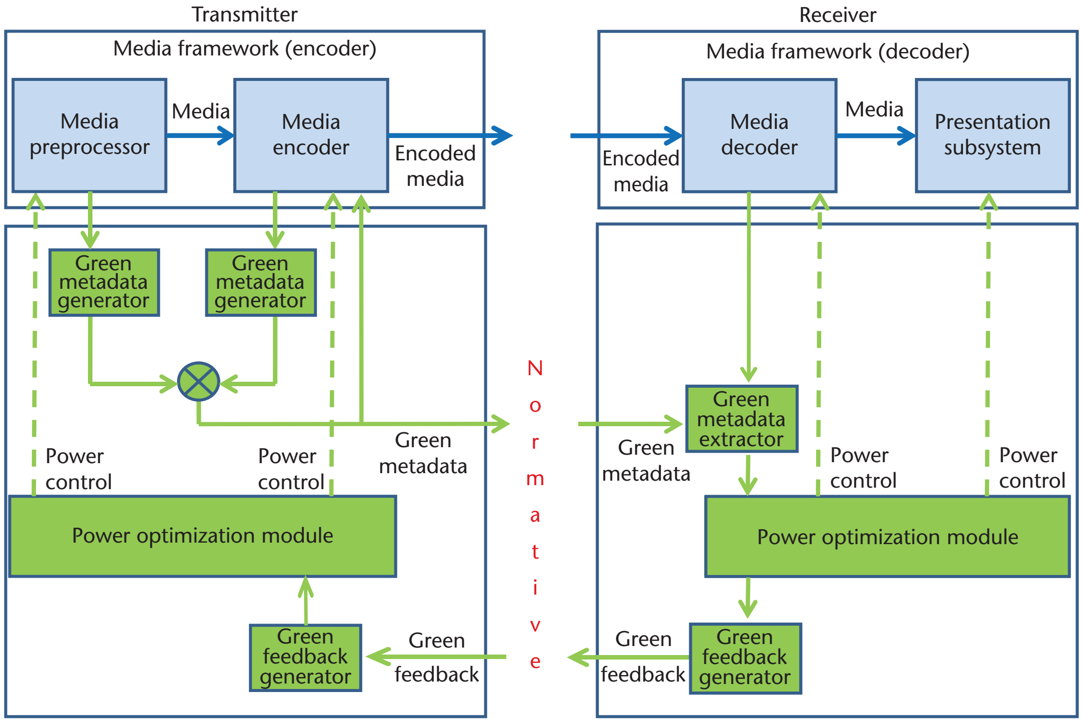

Figure 1. Functional architecture of a system that uses Green Metadata. The Green Metadata is generated during preprocessing and/or encoding at the transmitter.

been used to reduce display power consumption. However, when applied to video content, DA produces flicker artifacts. In this section, we explain how the metadata in the Green Metadata standard can be used to mitigate flicker and reduce power significantly when DA is applied to video content on mobile devices.

How Does DA Work?

The luminance of light perceived from a liquid crystal display (LCD) or an organic light-emitting diode (OLED) display is not only dependent on the transmittance of pixels (RGB values), but it is also proportional to the backlight intensity of the LCD panel (or the supply voltage of the OLED panel). Because the power consumption changes negligibly with variations of RGB values, DA can be utilized to reduce the power consumption considerably without sacrificing the quality by dimming the LCD backlight and simultaneously scaling up the RGB values proportional to the dimming level.

Figure 2 provides an example application of DA. Figure 2a shows a video frame displayed under the default full backlight. In Figure 2b, the LCD backlight is reduced by 26 percent. This reduces display power consumption by 26 percent, however, the image contrast is also reduced significantly. In Figure 2c, image quality is almost completely restored by using DA to scale up the RGB values. Depending on the amount of backlight reduction, some high RGB values may be saturated to the maximum level (255 for a 24-bit RGB image). For larger power savings, in Figure 2d the backlight is reduced by 65 percent. With this huge backlight reduction, simply scaling up the RGB values does not restore image quality, as seen in Figure 2e. Fortunately, contrast enhancement can be applied within a dynamic range that contains a majority of the RGB values.[3] Figure 2f shows significant improvement, compared with Figure 2e, after applying the aforementioned contrast enhancement.

How Does the Standard Use Metadata for DA?

Previous studies on DA focused exclusively on static images. Unfortunately, these techniques have limited applicability in realistic scenarios involving video content. First, when the techniques are applied to video sequences, flicker artifacts occur as a result of large backlight variations from dramatic changes in image characteristics at scene changes. A second limitation of the previous studies is that a significant latency (in milliseconds) exists between the instant when the backlight scaling control is applied and the instant when the backlight actually changes in response to the control. Therefore, the backlight values will not be

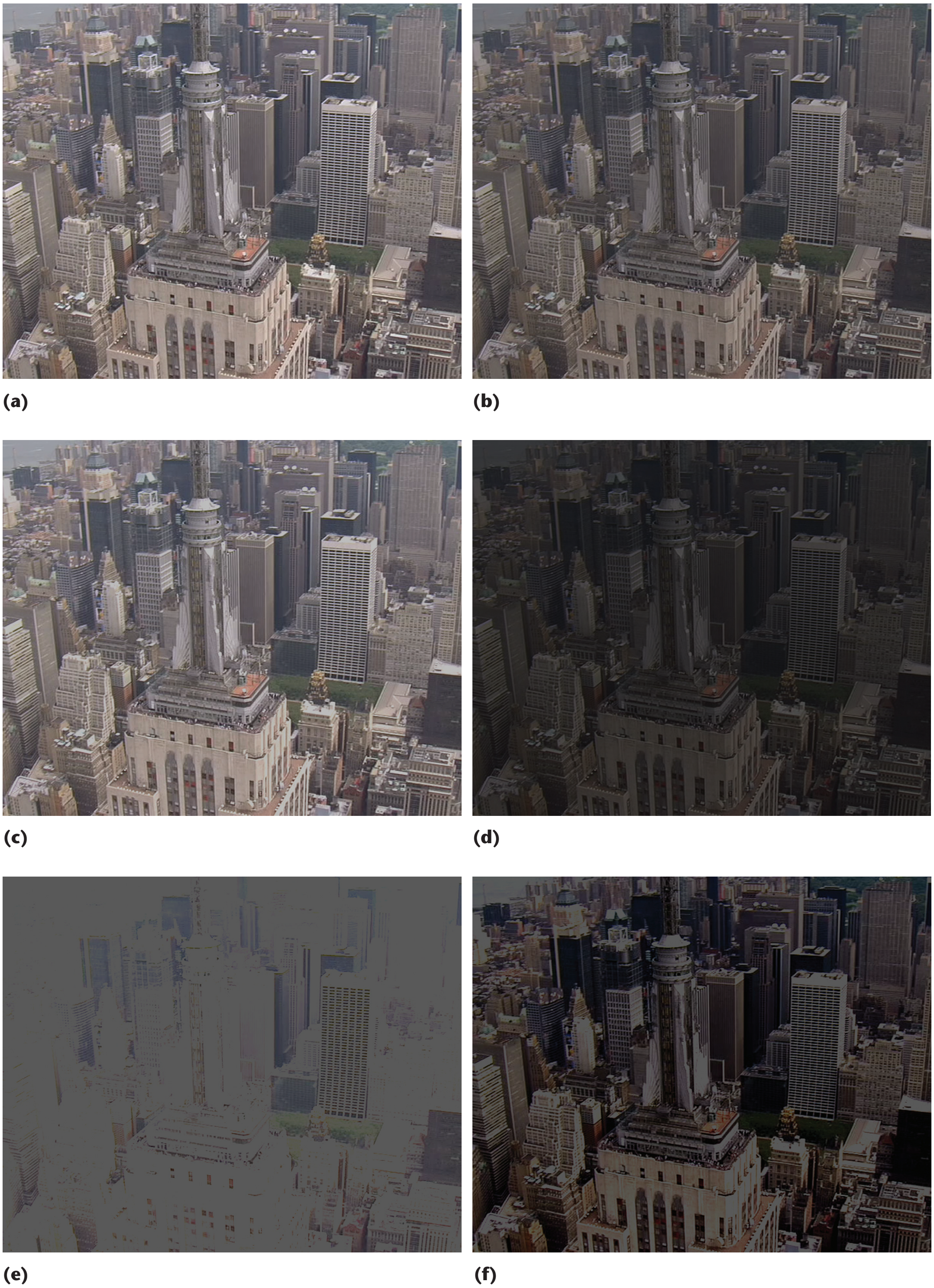

Figure 2. Example display adaptation (DA) application. (a) Video frame under full backlight. (b) Backlight is reduced by 26 percent, which results in contrast reduction. (c) Result of applying DA. (d) Backlight is severely reduced by 65 percent. (e) Result of DA without optimal dynamic range adjustment. (f) Result of DA with contrast enhancement using dynamic range adjustment.

synchronized with the displayed frames, and flickering is inevitable. A third practical limitation is that previous studies require DA to be applied successively to all video frames. However, existing displays cannot adjust the backlight for each video frame because the backlight cannot be adjusted at the video frame rate; the interval between backlight changes must be larger than the interframe interval.

The Green Metadata standard overcomes these limitations by signaling RGB statistics, associated quality-level indicators, and dynamic-range bounds as metadata. Backlight settings that will result in the indicated quality levels are derived from the RGB statistics. The dynamic-range bounds use contrast enhancement to improve perceived quality. For large (small) power reduction, the receiver's power-optimization module will select the backlight setting derived from RGB statistics associated with a lower (higher) quality level. The third limitation described earlier is overcome by aggregating the RGB statistics and associated quality levels over the interval between backlight changes. The second limitation is overcome by transmitting the metadata with sufficient lead time relative to the video frames so that the latency is not an issue. The first limitation is avoided by temporally smoothing the metadata so that large variations are absent.

To test DA power reduction, we used the Galaxy Tab 2 platform powered by the Monsoon power monitor, which replaces the tablet's battery and logs instantaneous power consumption of the tablet. Over a test set of 12 video sequences, we computed an average power reduction of 26.3 percent from metadata-assisted DA, without using contrast enhancement. When the metadata included the dynamic-range bounds that enable contrast enhancement, we measured up to 80 percent power reduction at acceptable quality levels. These reductions are relative to the power consumption of the device and are not limited to the display unit within the device. (Further details are available elsewhere.[5,6])

Decoder Power Reduction

The dynamic voltage/frequency scaling (DVFS) power-reduction technique has been extensively studied[7] and implemented in various systems. Intuitively, the dynamic power of a CMOS circuit is related linearly to the clock frequency f and quadratically to the supply voltage V, so that $P \propto f \cdot V^2$. Because V is a generalized power function of f, the power consumption may be expressed as

$$P \propto f^{2+a}. \tag{1}$$

Therefore, lowering the frequency reduces power consumption. For decoders that run on a CPU, we may apply Equation 1 to control the CPU frequency and thus reduce power. However, if the CPU frequency is lowered arbitrarily, then decoding of complex pictures may not complete within frame-rate deadlines. Consequently, the ability to accurately predict the complexity of upcoming pictures is critical.[8] Instead of using additional, complicated modules to predict the complexity, the Green Metadata standard adopts a new approach to this problem: metadata that indicates picture-decoding complexity is embedded in the bitstream and used by the receiver to set the CPU frequency at the lowest frequency that guarantees decoding completion within frame-rate deadlines and also provides the maximum power reduction.

Using Metadata to Indicate Decoding Complexity

Because the characteristics of video content usually change over time, picture decoding complexity also varies. Therefore, an efficient protocol for complexity signaling is to specify the period over which complexity is constant and to provide complexity metrics (CMs) that indicate the decoding complexity within that period. When content characteristics change, a new period is signaled along with accompanying CMs. The standard uses `period_type` to signal the type of upcoming period over which the CMs apply. The `period_type` is either a single picture, a group of pictures (GOP), or a time interval.

The total decoding complexity is decomposed[9] according to the complexity of six individual decoding modules (DMs): entropy decoding, dequantization and inverse transform, intra prediction, motion compensation, deblocking, and side-information preparation (such as memory copies). The complexity of each DM is a function of the complexity for a particular operation, which is typically expressed in cycles, as well as the number of times that the operation must be invoked in the specified period. The CM for a DM is the number of times that the DM's underlying operation is invoked within the specified period. The Green Metadata standard specifies a CM for each of the six DMs and embeds these CMs in supplemental enhancement information (SEI) messages within an H.264/AVC bitstream.

Experimental Results for CM-Assisted DVFS Decoding

To determine the power reduction from CM-assisted DVFS decoding, we used the Google Nexus 7 platform powered by the Monsoon power monitor. For on-platform bitstream playback, we used the software decoder in the RockPlayer application (www.rockplayer.com). Our measurements showed that CM-assisted DVFS decoding provides an average power reduction of 12.5 percent over a test set of 11 bitstreams on the Google Nexus 7 platform. This reduction is relative to the power consumption of the device and is not limited to the decoder module within the device. (Further details are available elsewhere.[5])

Encoder Power Reduction

An encoder can achieve power reduction by encoding alternate high- and low-quality segments in a segmented delivery mechanism such as DASH.[10,11] The power reduction occurs because low-complexity encoding mechanisms (fewer encoding modes, fewer reference pictures, smaller search ranges, and so forth) are used to produce the low-quality segments. A metric describing the quality of the last picture of each segment is delivered as metadata to the decoder. This section describes how cross-segment decoding (XSD)[12] can be used to improve the quality of the low-quality segments. An XSD-enabled decoder will utilize quality metrics contained in the high-quality segments (from high-complexity encoding) to enhance decoding of the low-quality segments (from low-complexity encoding), producing a visual experience with significantly higher QoE, but with reduced average encoding complexity (and therefore reduced encoding power consumption). The decoding complexity for the first picture in the low-quality segment is increased, while the decoding complexity for the other pictures remains the same as that for H.264/AVC-compliant decoders.

Cross Segment Decoding of Varying Quality Video

At the transition from a segment with higher video quality to a temporally neighboring segment that is encoded independently of the good quality segment and whose video quality is poorer, we term the last picture (in display order) in the higher quality segment a "good picture" (GP), the first instantaneous decoder refresh (IDR) picture of the poor quality segment the "start picture" (SP), and the output from the current algorithm the "fresh start" (FS). Note that the SP as an IDR picture was encoded without referencing the GP or any other frames in the higher quality segment.

The peak signal-to-noise ratio (PSNR) values of the GP and SP are embedded as Green Metadata in H.264/AVC SEI messages. When an XSD-enabled decoder receives this metadata, it will use the PSNR values to determine whether to apply an enhancement algorithm to enhance the SP. Specifically, the decoder can use information contained in the GP to improve the quality of the decoded SP to get FS, an improved reference frame for subsequent frames in the low quality segment. Once the decoder identifies the correspondences of the similarities between the two pictures, it can use the high-quality areas in the GP to improve the quality of the corresponding areas in the SP.

Depending on the motion level in spatial regions of the SP, the decoder may use different enhancement algorithms—one for relatively low motion areas, the other for the higher motion areas. For both algorithms, the decoder will look for matches between areas in the decoded GP and the SP.

Experimental Results for Encoder Power Savings with XSD

To demonstrate metadata-based power savings, we used the x264 AVC encoder installed on a Galaxy Tab 3. XSD modifications were applied to the JM 18.5 version of the AVC decoder. Power usage was measured using the Monsoon power monitor. Across 10 test sequences, an average of 12.1 percent power reduction was measured on the tablet. When compared with H.264/AVC-compliant decoding, the average PSNR, after XSD decoding, was improved by 0.2 dB. (Further details are available elsewhere.[13])

Energy-Efficient Media Selection

Over the past decade, delivery of multimedia over the Internet has become increasingly popular. The first commercial deployments were based on proprietary streaming platforms such as Apple's HTTP Live Streaming,[14] Microsoft's Smooth Streaming,[15] and Adobe's HTTP Dynamic Streaming (www.adobe.com/products/hds-dynamic-streaming.html). In 2011, MPEG standardized Dynamic Adaptive Streaming over HTTP (DASH),[10,11] the first open standard in this field. All these 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 decoder and the display of the client device.

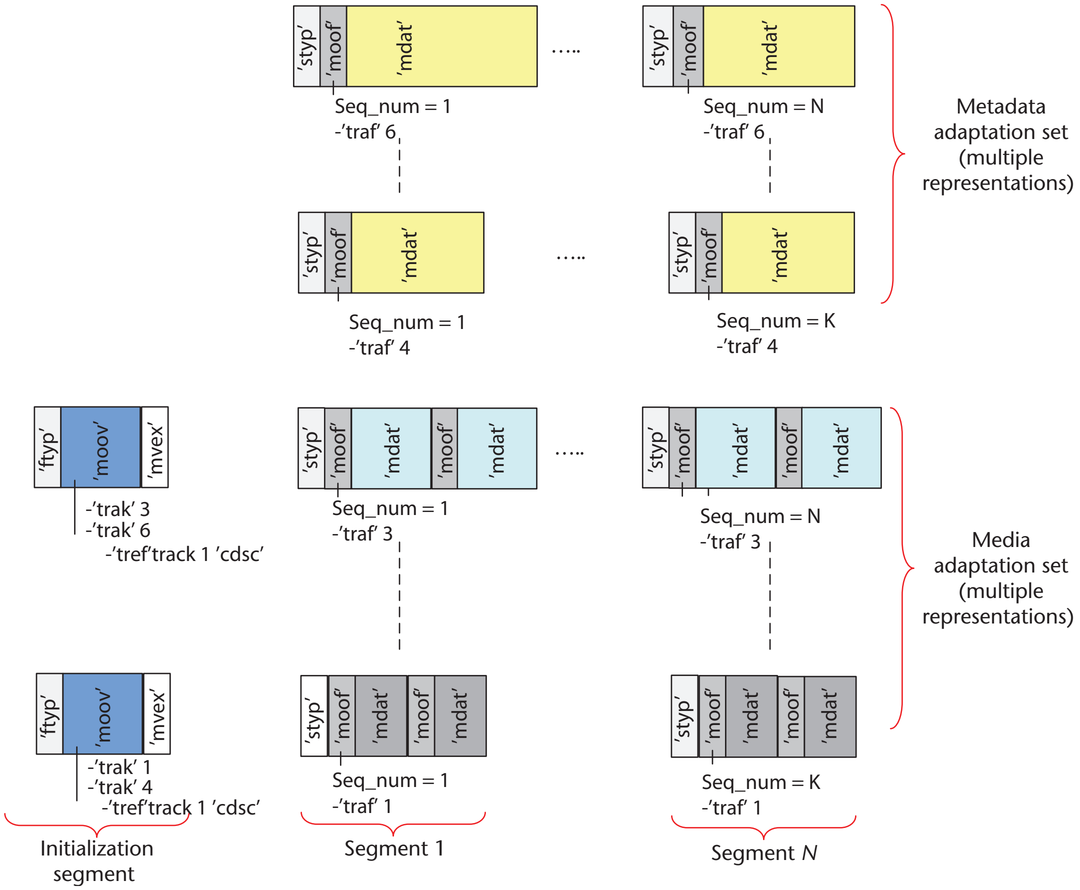

***Figure 3.** One metadata representation for one media representation. The segments are time aligned on segment boundaries.*

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

- Select a media representation that will require low power consumption. This leads to poor QoE because the lower resolution or lower bit-rate representations will be selected.
- 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. It can lead to visible changes of perceived quality, particularly for simple scenes when the resolution is increased.
- Use Green Metadata to select the most appropriate representation proactively 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 defines two types of metadata: decoder-power indication metadata and display-power indication metadata. These metadata are associated with available video representations for each segment in a media stream. The available metadata representations will be signaled in a specific adaptation set within the media presentation description (MPD) file, which describes a manifest of available content in DASH. The association of a metadata representation with a media representation is made in the MPD through the `@associationId` and `@associationType` attributes. A metadata segment and its associated media segment(s) are time aligned on segment boundaries, as shown in Figure 3.

The decoder-power indication metadata gives the potential decoder power saving of each available representation of a video segment. The potential power savings for a representation are indicated as two ratios. The first is the ratio of the representation's decoding-operations reduction (DOR) relative to the decoding operations (DO) of the most demanding representation in the current segment. The second is the ratio of the representation's DOR relative to the DO of the same representation in the previous segment. To determine the potential display power saving for a video segment, the display-power indication metadata consists of RGB statistics and the resulting quality levels as explained earlier.

The client device first determines the average acceptable power consumption based on its remaining battery life and the total duration of the video (parameter provided in the manifest file in case of on-demand content or requirements of duration expressed by the user). For each switching point, using the previously mentioned metadata and power consumption measured in the last segments, the player can define the best power-saving allocation strategy: the particular representation it needs to download and the appropriate RGB statistics that will provide an acceptable quality level and the desired power saving for that representation.

Additionally, if available in the video stream, complexity metrics can provide proactive DVFS processor control for optimal energy saving. RGB statistics can be used in addition to reduce display power consumption.

Conclusion

Because of the increasing popularity of video consumption on mobile devices, battery-powered systems that implement the functional architecture in Figure 1 are ubiquitous. Algorithmic, systemic, and modular optimizations are key to ensuring that such systems have low power consumption. However, huge additional power savings are achievable by considering the impact of the video content itself on the power consumption of various components within the functional architecture. The Green Metadata standard embraces this philosophy and specifies the abstraction of the power-consumption determinants of video content into a small amount of metadata that accompanies the video bitstream and that enables the additional power reductions to be achieved.

For display power reduction, the display-adaptation technology reduces power consumption by lowering the backlight (or voltage) control of the display. The power-consumption determinants are the RGB statistics, corresponding quality levels and dynamic-range bounds that enable backlight settings to be derived for the specified quality levels. These determinants are extracted from the video content at the transmitter, processed to avoid inducing flicker and then encapsulated in the bitstream as metadata.

For decoder power reduction, the dynamic voltage/frequency scaling technology reduces power consumption by lowering CPU frequency. The power-consumption determinants are the complexity metrics that indicate picture-decoding complexity so that the CPU frequency is set to the lowest frequency that guarantees decoding completion within frame-rate deadlines and also provides the maximum power reduction. These complexity metrics are extracted during video encoding and then encapsulated in the bitstream as metadata.

For encoder power reduction, the cross-segment decoding (XSD) technology enhances quality at the decoder after low-power encoding is used at the transmitter. The determinant is a quality metric that is used by the XSD-decoder to enhance the quality. The quality metric is computed at the transmitter and encapsulated in the bitstream as metadata.

For energy-efficient media selection, the power-consumption impact on the client's decoder and display is considered when media is selected at each switching point. The power-consumption determinants are display-power indicators and ratios of decoding-operations reduction relative to certain representations in the current and previous segments. These determinants are computed at the transmitter and signaled in a specific adaptation set within the Media Presentation Description file in DASH.

The Green Metadata standard is expected to become an international standard in February 2015. Given the huge demand for increased battery life on mobile devices, it is anticipated that an ecosystem will evolve to support the deployment of the standard. The ecosystem would consist of entities that generate, multiplex, distribute, and finally apply the metadata for power reduction. The French GreenVideo project (http://greenvideo.insa-rennes.fr) is currently considering deployment of the standard. **MM**

Acknowledgments

The work of Zhan Ma was performed when he was a senior standards engineer at Samsung Research America, Dallas. The work of Jiangtao Wen was partially supported by the National Science Foundation for Distinguished Young Scholars of China (grant 61125102), the State Key Program of National Natural Science of China (grant 61133008), and the National Significant Science and Technology Projects of China (grant 2012ZX01039-001-003-2).

References

1. Cisco Networks, "Visual Networking Index: Global Mobile Data Traffic Forecast Update 2013–2018," tech. report, 5 Feb. 2014; www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html.
2. *Information technology - MPEG Systems Technologies - Part 11: Energy-Efficient Media Consumption (Green Metadata), ISO/IEC JTC1/SC29/WG11 Final Draft International Standard 23001-11* (w14853), Oct. 2014.
3. W.-C. Cheng and M. Pedram, "Power Minimization in a Backlit TFTLCD Display by Concurrent Brightness and Contrast Scaling," *IEEE Trans. Consumer Electronics*, vol. 50, no. 1, 2004, pp. 25–32.
4. T.-H. Huang et al., "Enhancement of Backlight-Scaled Images," *IEEE Trans. Image Processing*, vol. 22, no. 12, 2013, pp. 4587–4597.
5. A. Tewari et al., "Samsung's Response to the Call for Proposals on Green MPEG," ISO/IEC JTC1/SC29/WG11, no. m30484, Aug. 2013.
6. E. Faramarzi et al., "Signaling and Metadata for Display Adaptation," ISO/IEC JTC1/SC29/WG11, no. m32480, Jan. 2014.
7. J. M. Rabaey, *Digital Integrated Circuits*, Prentice Hall, 1996.
8. Z. He et al., "Power-Rate-Distortion Analysis for Wireless Video Communication under Energy Constraints," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 15, no. 5, 2005, pp. 645–658.
9. Z. Ma, H. Hu, and Y. Wang, "On Complexity Modeling of H.264/AVC Video Decoding and Its Application for Energy Efficient Decoding," *IEEE Trans. Multimedia*, vol. 13, no. 16, 2011, pp. 1240–1255.
10. *Information technology Dynamic adaptive streaming over HTTP (DASH) Part 1: Media presentation description and segment formats, ISO/IEC Std. 23 009-1*, 2012.
11. I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming over the Internet," *IEEE Trans. Multimedia*, vol. 18, no. 4, 2011, pp. 62–67.
12. J. Wen et al., "Cross Segment Decoding of HEVC for Network Video Applications," *Proc. IEEE Data Compression Conf.*, 2013, pp. 1–8.
13. G. Wen et al., "Power Reduction through Cross-Segment Decoding," ISO/IEC JTC1/SC29/WG11, no. m32439, Jan. 2014.
14. R. Pantos, ed., "HTTP Live Streaming," Internet draft, Internet Eng. Task Force, 14 Oct. 2014; https://tools.ietf.org/html/draft-pantos-http-live-streaming.
15. A. Zambelli, Smooth Streaming Technical Overview," Internet Information Services, Microsoft, 31 Mar. 2009; www.iis.net/learn/media/on-demand-smooth-streaming/smooth-streaming-technical-overview

Felix C. Fernandes is a director of research and development at Samsung Research America. Contact him at felix.f@samsung.com.

Xavier Ducloux is an advanced studies program manager at Thomson Video Networks, France. Contact him at xavier.ducloux@thomson-networks.com.

Zhan Ma is a senior staff engineer at Huawei Technologies. Contact him at zhan.ma@outlook.com.

Esmaeil Faramarzi is a senior research engineer at Samsung Research America. Contact him at e.faramarzi@samsung.com.

Patrick Gendron is a technology officer at Thomson Video Networks, France. Contact him at patrick.gendron@thomson-networks.com.

Jiangtao Wen is a professor of computer science at Tsinghua University, China. Contact him at jtwen@tsinghua.edu.cn.
