CrowdDBS: A Crowdsourced Brightness Scaling Optimization for Display Energy Reduction in Mobile Video

Zhisheng Yan, *Member, IEEE*, Qian Liu, *Member, IEEE*, Tong Zhang, *Senior Member, IEEE*, and Chang Wen Chen, *Fellow, IEEE*

Abstract—Mobile display has become one of the most power-hungry component in mobile video viewing. Currently, mobile devices can reduce the display energy by performing dynamic brightness scaling (DBS) under the distortion constraint of video signals. We observe that there is a pitfall preventing current practice from systematic display energy reduction. In particular, existing objective DBS schemes lack direct connection to the subjective human perception on DBS-enabled videos, which is the key to achieving human-centered energy-experience optimization. To overcome this pitfall, we present CrowdDBS, a crowdsourced display energy reduction framework for mobile video viewing. CrowdDBS is empowered by a set of crowdsourcing studies that uncover the relationship between human perception and DBS frequency, magnitude, and temporal consistency, respectively. Motivated by the insights obtained from these studies, CrowdDBS employs a suit of designs and a DBS optimization framework to optimize the energy-experience tradeoff in mobile video viewing. Comprehensive experimental results and user evaluations under a variety of practical settings show that CrowdDBS can achieve 37% device energy reduction on average while guaranteeing satisfactory user experience in mobile video viewing.

Index Terms—Brightness Scaling, Crowdsourcing, Mobile Energy, Human Perception, Mobile Video.

1 INTRODUCTION

THANKS to the advancement of mobile web access and L online social networks, viewing videos on mobile devices has become a prevalent personal computing service. According to a recent study [1], video data will account for an impressive 75% of the global mobile data by 2019. Although the processing and storage capabilities of mobile devices have been rapidly expanding to meet the everincreasing video service, the limited battery power has long been a top gripe of mobile users. Among the powerconsuming components of mobile video viewing, mobile display energy is a key issue that needs to be addressed urgently. Not only have displays been recognized among the most power-hungry subsystems on mobile devices [2], [3], display has also been consuming energy unavoidably in video viewing. In particular, unlike networking components that can enter an idle/inactive state when not downloading the video data, energy dissipation of mobile display shall always occur as long as the video is playing.

Display energy is primarily consumed when illuminating the mobile screen. It is thus most effective to save the display energy by dimming the screen brightness. However, dimmed screen will drop the perceptual luminance of displayed pixels and may eventually degrade users' *quality of experience* (*QoE*) in mobile video viewing. Therefore, the key principle of current practices in display energy reduction is to dynamically scale the screen brightness while attempting to maintain a satisfactory user perception on video display. This technique is commonly termed as *dynamic brightness scaling* (*DBS*).

Despite the broad consensus in using DBS, we observe that our understanding of how to perform DBS is still limited. This might surprise many experts since DBS has been actively studied in design automation communities [4]-[8]. The reason behind this observation is that DBS introduces unique effects on the human perception of mobile videos whereas little is known regarding the optimal brightness under this new context. First, prior brightness scaling schemes for a single image [4], [5] cannot be directly applied to the multi-frame videos. Second, many exiting works [6], [7] were implemented and evaluated using programmable LCD modules, which cannot adequately reflect human perception on smartphones that employ larger screen size and higher pixel-density. More importantly, existing DBS schemes using objective viewing distortion metrics, e.g., linear relation between perceptual luminance and brightness levels [8], [10], lack direct resonance with the mobile viewing experience. Thereby, it is imperative to design an effective DBS scheme with human-in-the-loop modeling that directly resonates with the QoE of DBS on modern mobile devices.

The objective of this research is to explore the human perception when DBS is applied on mobile display and model the user experience to characterize this special humansystem interface. By exploiting the perception models of

Z. Yan is with Department of Computer Science, Georgia State University, Atlanta, GA 30303, e-mail: zyan@gsu.edu.

C. W. Chen is with Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, NY 14260, e-mail: chencw@buffalo.edu.

Q. Liu is with Department of Computer Science and Technology, Dalian University of Technology, Dalian, Liaoning 116024, e-mail: qianliu@dlut.edu.cn.

[•] T. Zhang is with Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, e-mail: tzhang@ecse.rpi.edu.

DBS, we ultimately aim at minimizing the display energy consumption while providing a satisfactory video viewing experience. To achieve this objective, we are facing several technical challenges.

- Finding a *perceptually suitable scaling frequency*: Less brightness scaling brings more stable viewing experience, but will lead to much smaller space for manipulating the brightness dynamics to achieve desired energy reduction [8]. When to apply the scaling also matters because varying brightness abruptly would degrade user experience.
- Determining a *perceptually optimal scaling magnitude*: Deciding the brightness for a given segment of video frames is challenging because the inherent relationship between user experience and brightness level has not yet been understood clearly. Furthermore, human visual perception of a certain brightness also depends on the displayed video content.
- Applying *temporally consistent* DBS across segments: Human eyes are extremely sensitive to temporally unsmooth videos. Hence, annoying flicker can be introduced if the brightness of each segment is scaled independently without addressing inter-segment variation.

To tackle these challenges, we present *CrowdDBS*, a display energy reduction framework that integrates the optimal DBS service using crowdsourced data. To understand human perception of DBS, we carry out the first crowdsourcing studies in order to explore the hidden connection between user experience and scaling frequency, magnitude, and temporal consistency, respectively, through a customized crowdsourcing App. Building on the crowdsourced data and statistical analysis, we notice that a scaling frequency aligned with video shot boundaries is on average 11 times more acceptable by users than other alternative schemes. Therefore, we propose a new threshold based video shot detection algorithm to pinpoint the scaling points at shot boundaries. Furthermore, we discover a content-luminancedependent logistic relationship (instead of linear) between user acceptability and brightness levels. We then construct a new user experience model using logistic regression in order to determine the best scaling magnitude for a given video shot. Moreover, we find that users are usually sensitive to down-scaled brightness variations rather than the upscaled ones. Accordingly, a Weber's law based model is developed to restrict the brightness down-scaling between two consecutive shots.

To optimize the human-in-the-loop designs in CrowdDBS, we formulate and analyze a DBS optimization problem by exploiting the abundant real user measurements. We seek the optimal brightness levels for each video shot of a given video in order to minimize the display energy subject to user acceptability and temporal consistency. We then propose an efficient algorithm that is proven to be optimal in order to solve the optimization problem. Finally, combining all these designs and algorithms, we integrate CrowdDBS into the use cases of mobile video streaming and offline video viewing to illustrate its broad applicability.

We prototype CrowdDBS in commercial smartphones. We validate CrowdDBS designs and algorithms using realworld experiments. We evaluate the design components of CrowdDBS, as well as conducting system-level experiments under a variety of practical settings including different application scenarios, video contents, mobile devices, and video shot lengths. The results show that CrowdDBS can save 37% device energy on average while guaranteeing satisfactory user experience in mobile video viewing.

To summarize, the contributions of the proposed research include:

- A suite of crowdsourcing studies to explore the relationship between human perception and scaling frequency, magnitude and temporal consistency (Section 3).
- A DBS optimization framework using a set of designs motivated by the studies to minimize the display energy with satisfactory user experience (Section 3-4).
- A practical demonstration of the desired performance achieved by the proposed display energy reduction with crowdsourced DBS optimization (Section 5-6).

2 RELATED WORK

2.1 Dynamic Brightness Scaling

Organic light-emitting diode (OLED) and liquid crystal display (LCD) are two mainstream display technologies for modern smartphones that feature different energy consumption principles. In this paper, we focus on screen brightness scaling, a pragmatic display energy reduction strategy that can be employed by both LCD and OLED devices via dimming the backlight [11] and decreasing the luminance of displayed image [3], [12], [13], respectively.

The operating procedure of existing DBS schemes mainly consists of two steps. By assuming a linear relationship between human perception and brightness level, the perceptual luminance of a brightness-scaled video can be computed first. The brightness levels are then scaled down continuously until a signal distortion threshold between the scaled and original videos is reached.

In particular, Cheng *et al.* aimed at dimming the brightness while simultaneously retaining the image contrast [4]. Iranli *et al.* pre-converted the original image using histogram equalization such that the image fidelity can be preserved after dimming [5]. These single-image based strategies demonstrate the principle of brightness scaling, but are not directly applicable to multi-frame videos.

Several video DBS schemes were also developed by replacing image contrast metric with video-related distortion metrics, such as structural similarity (SSIM) [7] and peak signal-to-noise ratio (PSNR) [10], [14]. Recently, the SSIMbased DBS strategies was optimized via a dynamic programming algorithm [8]. The authors used a cloud-assisted architecture to further reduce mobile streaming energy by removing the on-device distortion computation. Similarly, to expand display energy saving, the GPU on mobile devices was utilized to perform the distortion analysis in [9]. However, these works all rely on objective distortion formulations using signal analysis at individual device and therefore lack direct connection with user experience. Instead, we extend the energy-experience tradeoff by probing the subjective perception of numerous mobile users. We investigate the human perception of DBS-enabled videos on modern smartphones through crowdsourcing studies. We aim at leveraging the crowdsourced insights to understand the relationship among QoE, backlight and video content, and thus to maximize the energy saving of video viewing systems.

2.2 Crowdsourcing and Video Quality

Crowdsourcing has become a popular utility to obtain services and opinions from a group of people via online or offline study. Since videos directly render the real-world scenes perceived by human eyes, it is natural to make use of crowdsourcing to understand video perception. Typical research efforts include crowdsourcing for video quality [15], video annotation [16] and user viewing pattern [17]. Crowdsourced results present higher correlation to human perception than the traditional objective approaches. For example, they capture user experience more effectively than those objective video quality metrics that utilizes low-level signal features, e.g., PSNR and SSIM [18].

A more common methodology for video quality crowdsourcing is to model QoE as mean opinion scores via user data. Such efforts were made in a variety of use cases, including adaptive HTTP streaming [19] and voice-over-IP [20]. One potential issue for such score-based tests is that participants may be overburdened by the multiple levels of scores and may struggle with making a correct decision [21].

Recently, binary-choice assessment was suggested to evaluate the acceptability of videos [18], [21], [22]. The lowest acceptable video quality for a pleasant viewing can be found using Method of Limits [23], where users view a series of videos varied in successive quality steps with descending or ascending order and submit their binary acceptability choice.

All these existing works explore how user experience is impacted by video encoding/transmission parameters, e.g, resolution and delay, in order to enhance the networking performance of video viewing/streaming systems. The objective of this paper, however, is fundamentally different in that we probe into the new QoE space of brightness scaling in order to enhance the device energy saving.

3 CROWDSOURCING BASED DBS DESIGNS

In order to replace the objective video signal based DBS, we introduce the proposed DBS designs by crowdsourcing human perception over a network of users in this section.

3.1 Crowdsourcing Studies Setup

We first present the overall configuration of the three crowdsourcing studies.

Test Devices. We use three test devices, i.e., LG Optimus G Pro, Google Nexus 4, and Xiaomi Mi 2. G Pro and Nexus 4 are employed by default except that in the video consistency study G Pro and Mi 2 are used. The display size and resolution of G Pro, Nexus 4, and Mi 2 are 5.5 inch/1080x1920, 4.7 inch/768x1280, 4.3 inch/720x1280, respectively.

Participants. A total of 50 users (32 males and 18 females, age 20 to 39) with normal or corrected vision were recruited via university email list and online forum. The number of participants meets the requirement of ITU video quality evaluation [24] and is also comparable with existing crowdsourcing studies [16]. In all studies, users are equally split into two groups and each group uses a different mobile device.

Incentives. We used monetary incentives to recruit users to complete the studies on DBS scaling frequency, magnitude and temporal consistency. Such a mechanism was essentially a platform based scheme where we provided a fixed reward to a user finishing a task. However, different users may provide different quality of information. An incentive algorithm that maximizes the overall quality of information while meeting the reward budget (or minimum user number) can be designed as the future work to enhance the user studies.

Video Sources. We collect 30 videos as the sources. These sources will be segmented into test clips for the crowdsourcing studies. Since we concentrate on the perception of DBS, we encode all video sources identically at 23 quantization parameter, 24 frames/second, 1280x720 resolution, and H.264/AVC baseline profile. The audio channel is encoded at 44KHz and >180Kbps to ensure that participants can understand the content and obtain the most realistic viewing.

Assessment Method. Considering the less burden during its decision-making process, binary-choice test is adopted as in prior works [18], [21], [22]. We ask users to recognize whether or not a test clip is acceptable. The acceptable quality is defined as the video quality with tiny/no change of screen brightness that one would have a pleasant viewing experience in everyday life.

Test Environment. Unlike traditional subjective tests performed in strictly controlled lab environment, our crowdsourcing studies instruct the participants to view the videos comfortably without any time or pose restrictions. Another factor that may impact human perception of luminance is the ambient light. Since people spend majority of their video viewing time indoor, we focus on the indoor residential and working lighting (100 to 1000 lux) [25]. As reported in a recent study [26], more than 81% of users watch mobile video at home or working place. Therefore, it is clear that a greater bulk of mobile video viewing occurs indoor.

Test App. We implement an open-source¹ Android App for the crowdsourcing tests. After providing the personal information (Fig. 1a), participants will view a test clip that has several versions differed by a certain stimulus (Fig. 1b). Upon a version is clicked, the ambient light will be measured via the phone's light sensor. If the lighting is not suitable, a warning window will pop up and instruct the participant to change to a darker or brighter environment.

To realize DBS, a control thread is created by the player to dynamically vary the brightness via Android API. It also guarantees that the timing of DBS is synchronized with that of video/audio codec to support fast forward, pause, etc. If allowed by test protocols, users can double

1. https://github.com/yanlove/QoEPlayer



(c) Pop-up Dialogue

(d) Opinion Form

Fig. 1. Screenshots of the crowdsourcing App.

tap the screen during the playback to submit their decision (Fig. 1c). Alternatively, an opinion form (Fig. 1d) will be always shown when the playback ends.

Thanks to such an automatic process, more immersive viewing experience and more reliable human perception can be obtained. We emphasize that the test App can be modified and ported to any human perception based mobile video studies.

3.2 Addressing DBS Frequency

In this section, we obtain the perceptually suitable scaling frequency for CrowdDBS. Essentially, we need to find the optimal segmentation of a given video, where the brightness is scaled at each segment boundary. We first identify two important requirements. First, the brightness should be scaled at those video frames where users pay less attention in order to ensure a visually appealing playback. Second, the video should be split into more segments, which would allow more space to manipulate the brightness dynamics and thus save more power [8]. For example, if we mix two segments whose lowest brightness to maintain user experience are b_1 and b_2 ($b_1 \neq b_2$), the lowest acceptable brightness for the combined segment would be max{ b_1, b_2 }. As the display power is a nondecreasing function of b, this will result in the increase of power, i.e., $P(b_1) + P(b_2) < 2P(\max{b_1, b_2})$.

3.2.1 Crowdsourcing Study

Ideally, we should inspect all possible segmentation options of a video and evaluate how each DBS frequency impacts user experience. However, this is infeasible since there would be too many segmentation choices to go through, e.g., a 1-second video will produce tens of millions options. Hence, we propose to compare multiple frequency strategies and infer a suboptimal yet acceptable solution. To the best of our knowledge, there are only two existing methods to decide DBS frequency. One is to scale the brightness every constant number of frames (denoted by *Const*) [8]. The other scheme [27] segments the video such that the variance of frame luminance within one segment is less than a threshold (denoted by *GoS*).

In addition, we propose a new method that uses a *video shot* based segmentation to obtain the DBS frequency. A shot is defined as a set of consecutive video frames for a scene that were captured by the camera for an uninterrupted period of time. Once the camera stops continuous capturing and changes to the other angle/scene, a new video shot begins. The intuition behind this proposal is that there are always sudden transitions of video contents at shot boundaries. Therefore, users are expected to pay much less attention to these boundary frames than to the within-shot frames showing the actual content. Note that the video scene at semantic level is a similar concept. We choose shots instead of scenes because a scene usually includes multiple shots and thereby provides less space for energy reduction.

Procedure and Results. By applying the three candidate DBS frequencies, we dynamically scale the brightness of 4 test clips (2 movie clips and 2 full trailers). We inherit the default values of constant frequency (10 frames) and variance threshold (40) as in [8] and [27], respectively. To enforce the shot-based frequency, a new shot detection algorithm (see Section 3.2.2) is utilized and the falsely detected or missed shots are manually corrected. For all three frequency schemes, the brightness level of each segment *i* is configured by the example mapping function in [27]. That is, if the content of a segment is darker, its brightness can be lower and vice versa:

$$b_{i} = \begin{cases} \max(b_{i-1} - \Delta b, 0) & Y_{i} < Y_{i-1} \\ b_{i-1} & Y_{i} = Y_{i-1} \\ \min(b_{i-1} + \Delta b, 1) & Y_{i} > Y_{i-1} \end{cases}$$
(1)

where Y_i is the average luminance of segment *i* over all frames and pixels, and $\Delta b = 0.2$ is the varying step of brightness. This Δb is selected to prevent too little or too much variations that might cause all or none of the frequencies acceptable. Note that this brightness setting may not be optimal. We choose this simple mapping scheme only to better study the impacts of scaling frequency. Users are then required to view the three frequency versions randomly. They are asked to identify an unacceptable quality during the video playback, but wait until the end of playback to verify an acceptable quality.

In binary-choice assessment [18], [21], [26], human perception towards a video can be modeled as user acceptability A, i.e.,

$$\mathcal{A} = \frac{N_{acc}}{N_{tot}} \tag{2}$$

where N_{acc} is the number of votes for acceptable quality and N_{tot} is the total number of votes. We show the user acceptability under different scaling frequencies in Fig. 2. It can be seen that user experience is clearly impacted by scaling frequency. *Chi-square test* between frequency strategies and acceptability, $\chi^2 = 167.008 \ df = 2 \ p < 0.001$, further sup-



Fig. 2. User experience under different scaling frequencies on G Pro (top) and Nexus 4 (bottom).

ports that human perception of different scaling frequencies significantly differs from each other in a statistical sense.

We also observe that the video shot based DBS frequency achieves the highest user acceptability in all test cases. It obtains an average acceptability of 88.75%, which is 1.82 times and 20.3 times higher than GoS and Const, respectively. This can be attributed to the fact that users pay most attentions to the essential parts of a video, e.g., people, stories, and objects. Subconsciously, they do not quite care about the shot transition. Thus brightness switch when presenting those essential contents are much more annoying than that at shot boundaries. We can now conclude that shot-based DBS frequency provides a simple yet practical solution to this challenging problem. Furthermore, experience are found to be content dependent, which implies that the optimal shot brightness should be based on the content features.

3.2.2 A Shot Detection Algorithm for DBS

We proceed by proposing a new shot detection algorithm that is tailored for the human perception driven scaling frequency in CrowdDBS. The general idea of shot detection includes two phases [28]: calculating discontinuity metrics for consecutive frames and detecting the shot boundaries using the discontinuity values.

In CrowdDBS, we propose to adopt a luminance based discontinuity metric for DBS. This is because luminance metric is robust to missed detection in the context of DBS. Specifically, missed detection only lies between two adjacent shots with very similar luminance. However, this is not risky as we would have assigned the same brightness level to both shots even if we detected them. Therefore, our detection goal is to pinpoint the hard-cut shots that present abrupt luminance changes and no overlapped content in two consecutive frames. It is not necessary to detect nonhard-cut gradually-transited shots like dissolves, where the consecutive frames contain the content from both neighbor shots.

In particular, we employ Earth Mover's Distance (EMD) [29] as the frame discontinuing metric of the proposed algorithm. EMD is defined as the minimum cost to transform one luminance histogram into the other. We choose EMD since it is commonly used in image retrieval to differentiate



Fig. 3. EMD of "sniper" (first 1600 frames).

images. More importantly, EMD is a cross-bin feature that considers both inter-bin distance and bin height. Unlike bin-to-bin features adopted in traditional shot detection algorithms, EMD effectively captures the difference of luminance perception, improving the detection in CrowdDBS. Mathematically, the EMD between histogram P and Q is expressed as:

$$EMD(P,Q) = (\min_{f_{xy}} \sum_{x,y} f_{xy} D_{xy}) / (\sum_{x,y} f_{xy})$$

s.t.
$$f_{xy} \ge 0, \sum_{y} f_{xy} \le P_x, \sum_{x} f_{xy} \le Q_y$$

$$\sum_{x,y} f_{xy} = \min(\sum_{x} P_x, \sum_{y} Q_y)$$
(3)

where f_{xy} is the intensity amount moved from *x*th bin to *y*th bin and D_{xy} is the distance between bin *x* and bin *y*.

To detect the shot boundaries, we must find the frames with significant discontinuity. We exploit both global thresholding via absolute EMD and local thresholding via relative EMD. The motivation is that there is a wide range of EMD values at true hard-cut positions, as exemplified in Fig. 3. If only a global threshold θ is used (as in traditional algorithms), those non-hard-cut positions with EMD greater than θ would be falsely detected. By including relative threshold η , we can avoid these false detections. Furthermore, since per-frame scaling usually cannot be supported due to hardware limit, the shot length should have a minimum bound L_{\min} , which is found to be five frames based on the measurement of the three test devices. Note that this hardware limit will not necessarily affect energy performance as most shots last more than 1 second. We summarize the proposed shot detection in Algorithm 1. Specifically,

Algorithm 1 Proposed Shot Detection Algorithm							
1:	$\mathcal{S} \leftarrow 0, EM$	$D_{pre} \leftarrow EMD(0,1)$					
2:	for $k \leftarrow 1; k$	$c < LastFrame; k \leftarrow k + 1 $ do					
3:	if EML	$P(k, k+1) > \theta \& \frac{EMD(k, k+1)}{EMD_{nre}} > \eta$ then					
4:	$\mathbf{if}\;k$ -	$+1 - s_{pre} \ge L_{\min}$ then					
5:	S	$\leftarrow \mathcal{S} \cup \{k+1\}, s_{pre} \leftarrow k+1$					
6:	end i	f					
7:	end if						
8:	EMD_{pr}	$_{e} \leftarrow EMD(k, k+1)$					
9:	end for						
10:	return \mathcal{S}	$\triangleright S$: set of start frames for shots					

the EMDs between frame k and its neighbor frames are computed first. If EMD(k, k+1) passes the global threshold and is sufficiently larger than EMD(k-1, k), as well as the shot length lower-bounded by L_{\min} , we can conclude that a new video shot starts from frame k + 1.



Fig. 4. User experience under different brightness levels on G Pro (left) and Nexus 4 (right).

3.3 Modeling DBS Magnitude

In this section, we explore how human perception of a given shot is impacted by DBS magnitude, i.e., the brightness level. Based on crowdsourced data, we model the perception-brightness relationship, which will be applied into the CrowdDBS system later.

3.3.1 Crowdsourcing Study

As content luminance affects human perception of brightness, we must understand the acceptability of video shots with different luminance features under different brightness levels. Considering the advantages of EMD in characterizing content luminance in Section 3.2, we adopt EMD as the luminance feature of a shot. In order to compute EMD, a histogram of a shot needs to be extracted. Median histogram is a histogram descriptor for a set of images that has been broadly utilized in video research [30]. It is derived by computing the median bin height of each bin over all the frame histograms. The reason for selecting this descriptor is that it can eliminate the outlier frames within the shot effectively [30] and thus characterize the perceptual luminance accurately. Finally, the luminance feature of a shot EMD_{shot} is defined as the EMD distance between the shot's median histogram and the histogram of a black (zerointensity) image. A greater EMD_{shot} implies a brighter content.

Procedure and Results. For the user study, we first discretize the range of EMD_{shot} to represent different luminance levels of contents. After computing EMD_{shot} for all shots in all the source videos, we observe that the value ranges from 16 to 185.3 (2041 values in total). We divide EMD_{shot} into 5 categories (category 5 darkest) at a step of 34 and accordingly prepare 5 single-shot test clips. In addition, we vary the brightness levels b from 0.95 to 0.05 at a step of 0.1, where 1.0 represents the full brightness. Note that the brightness is scaled only once at the beginning of the shot. Participants start with viewing a full-brightness version as the reference and then 10 scaled versions are presented from either the highest or the lowest brightness using Method of Limit as in [18], [21]. Participants identify whether or not each version is acceptable, where the "acceptable quality" becomes the fixed brightness at which participants would perceive no/tiny difference compared with the full-brightness clip and would enjoy everyday. Participants are allowed to submit acceptable/unacceptable decisions during the playback.

Eventually, a total of 2500 binary data points are collected in this study. To grasp a clear understanding of the data, we illustrate the binary data as acceptability in Fig. 4. It is clear that there is a nonlinear relationship between user experience and brightness, indicating that past assumptions about linearity, e.g., [8], [10], is not accurate. In fact, a sigmoid curve ("S" shape curve) is observed for all luminance categories. Furthermore, it is interesting to see that the sigmoid curve is shifted horizontally as EMD_{shot} changes. For example, under the same brightness, users express higher acceptability towards a darker content than a brighter content.

3.3.2 Logistic Regression Analysis

To predict user experience under a brightness level and a particular video's content luminance feature, we propose to employ logistic regression for the prediction model training. Logistic regression analyzes binary data whose probability of being positive is a sigmoid curve, which perfectly matches our data shown in Fig. 4. Besides, logistic regression has been recommended for QoE modeling by ITU-T [31]. The logistic function can be expressed as:

$$F(\vec{x}) = \frac{1}{1 + \exp^{-(\alpha + \beta x_1 + \lambda x_2 + \delta x_3 + \dots)}}$$
(4)

where α , β , λ , \cdots are the coefficients and $\vec{x} = (x_1, x_2, \cdots)$ are the predictors. In this study, the predictors are the three stimuli, i.e., brightness level of a shot, shot EMD (content feature), and device type while the coefficients are determined by maximum likelihood estimation. We also exclude outlier points with studentized residuals less than -2 or greater than 2, which follows the standard logistic regression in [32].

We observe in the modeling process that the device type is not a significant predictor although other predictors pass the significance tests. Specifically, Wald test of device term (z = -1.313, p = 0.189) indicates that device type does not significantly contribute to the model fit in a statistical sense. A likelihood ratio test between the two models that include and exclude device type results in $\chi^2 = 1.734 \ df = 1 \ p = 0.187$, which further validates that the improvement of model fit by incorporating device type is not statistically significant. This is interesting since it is opposed to traditional understanding of human perception in video encoding, where a large screen relatively enlarges the coding artifacts and indeed deteriorates user experience. In the case of DBS, however, people would always have an absolute perception of the brightness regardless of screen size or pixel density. Since it is preferable to settle a simple model, we remodel the data by dropping the device term and only including the other two predictors. The modeling results is shown in Table 1. Note that the predictors of brightness and shot EMD both pass the Wald test and likelihood ratio test and are statistically significant. Finally, we summarize the user acceptability prediction model as

$$\mathcal{A} = \frac{1}{1 + \exp^{-(\alpha + \beta E M D_{shot} + \lambda b)}} \tag{5}$$

where the coefficients are listed in Table 1. Consequently, when we have a new test video, we will compute its EMD_{shot} value and replace the EMD_{shot} predictor by this value to predict the user acceptability.

TABLE 1 Model Coefficients

	Term	Estimate	Wald Test	Likehood Ratio Test		
ſ	Constant	-9.208	z = -13.69	NA		
	EMD_{shot}	-0.081	z = -13.62	$\chi^2 = 2526.5 df = 1964$		
	b	28.604	z = 15.09	$\chi^2 = 454.1 df = 1963$		
*All tests are significant at $p < 0.001$.						

3.4 Smoothing DBS Inconsistency

We now proceed to study if continuously playing brightness-scaled shots using the proposed scaling magnitude would incur the inconsistent variations, and how to address such annoying flicker effects in CrowdDBS.

3.4.1 Crowdsourcing Study

We proceed by evaluating the human perception of viewing two consecutive shots. We aim at understanding brightness variation at this minimum unit in order to extend the treatment of temporal consistency to the entire video. Specifically, we identify whether or not the brightness level of the second shot is perceptually suitable when given the brightness of the first shot.

Procedure and Results. A set of two-shot test clips is prepared such that the luminance feature EMD_{shot} of both shots span from category 1 to 5. We exclude the five clips whose luminance of both shots belongs to the same category. The reason is that, under the same acceptability threshold, the proposed brightness of these two shots would be very similar based on (5) and will definitely show smooth playback. This way, we produce a total of 20 test clips. Furthermore, we fix b_1 of the first shot as the perceptually lowest brightness $b_{1,opt}$ using (5) and $\mathcal{A} = 0.9$. The test brightness of the second shot b_2 for down-scaled clips $(b_2 < b_1)$ is gradually decreased from b_1 at a step of 0.1 until reaching the perceptually lowest brightness $b_{2,opt}$, i.e., $b_2 = \{b_1 - 0.1, b_1 - 0.2, \cdots, b_{2,opt}\}$. We do not study the cases when $b_2 \ < \ b_{2,opt}$ because the brightness of those second shot would be too dark to guarantee user experience. For instance, suppose the test clip has $b_{1,opt} = 0.4$ and $b_{2,opt} = 0.7$, and the second shot is scaled to 0.4. Even though the brightness variation is smooth, the second shot presents unacceptable quality because its brightness was already below 0.7 at the first place. Similarly, only one version is prepared for up-scaled test clips, i.e., $b_2 = b_{2,opt}$. Following the same test protocol in frequency study, a total of 1500 binary data points from 30 two-shot versions (20 down-scaled and 10 up-scaled) have been collected.

We discover an interesting distinction between the human perception of down-scaled and up-scaled clips. It can be seen from Fig. 5 that all up-scaled clips are generally acceptable. This result is compatible with previous study on human perception of video bitrate variation, where lowquality segments followed by high-quality segments was more visually appealing to users [33]. It might be explained as that users subconsciously believe brighter video is better and thus enjoy the up-scaled clips. Therefore, we conclude that $b_{1,opt}$ and $b_{2,opt}$ can be directly adopted for up-scaled shots. On the other hand, we find that the user experience of down-scaled clips degrades as the brightness variation enlarges. Furthermore, under the same level of variation,



Fig. 5. User experience of different patterns for up-scaled variations (top) and down-scaled variations (bottom).

human perception can still differ for some clips. Given these complex connections, it is imperative to address downscaled variation carefully.

3.4.2 Modeling DBS Variation

After scrutinizing the crowdsourced results for down-scaled clips, the human perception pattern of brightness variation can be summarized as follows. Under the same level of brightness variation, user acceptability degrades as the brightness of the first shot b_1 decreases. For example, due to the lower luminance of the first shot, the clip with EMD_{shot} switching from category 2 to 4 (denoted by 2->4) has a lower first-shot brightness than the clip with 1->3. Under the same brightness variation $\Delta b = 0.2$, clip 2->4 thereby presents a lower user acceptability. Similarly, we can observe a phenomenon that, to maintain a certain acceptability, the maximum allowed Δb for a clip decreases as b_1 decreases.

These effects exactly match the nonlinearity of human eyes. That is, the luminance perception of human eyes is not determined by the nominal change in light energy, rather by the stimulus change relative to its initial level. Therefore, the perception is not simply depended on Δb . Instead, it is dictated by the whether or not Δb is significant compared to b_1 . Moreover, we observe that as long as the ratio between Δb and b_1 is less than a threshold, the experience is acceptable regardless of which video the user views.

Building on the above analysis, such effects of human eyes can be modeled with the classic *Weber's law* [34], which states that the maximum unnoticeable change of stimulus is a constant ratio of the initial stimulus. Thus the maximum down-scaling change of brightness for the second shot can be written as

$$\Delta b \le C b_1 \tag{6}$$

where C ($0 \le C < 1$) is the constant ratio. Since this relationship is independent of the actual video under viewing, it can be applied to new test videos.

4 DBS OPTIMIZATION IN CROWDDBS

Putting together the abundant crowdsourcing findings and designs in the previous Section, we now convert the real measurement results into theoretical formulation and optimization for dynamic brightness adaptation strategy of CrowdDBS.

4.1 Problem Formulation

The objective of optimization is to seek the dynamic brightness levels of a video to minimize display energy while meeting the acceptability requirement.

We first discuss the display power model. Based on our own measurements, the display power is a nondecreasing and nonlinear function of the brightness b, which presents a general power law model induced by the gamma correction within the display hardware [3]. This finding is consistent with previous measurements [2], [3]. Therefore, the display power P_{disp} can be expressed as

$$P_{disp} = \varphi b^{\gamma} \tag{7}$$

where $\varphi > 0$ and $\gamma > 1$ are device-dependent constants. Note that the display power of OLED display is determined by both brightness *b* and pixel luminance indicator *l*. Thus we have $P_{disp} = \kappa (b \cdot l)^{\gamma} = \varphi b^{\gamma}$ for OLED devices, where constant φ is now both device and content dependent.

Recall that the acceptable scaling frequency of a given video can be derived using the proposed shot detection algorithm. This process is independent of the brightness optimization and can be completed in advance because shot boundaries are only decided by frame pixel values rather than by brightness setting or display energy. Therefore, given the scaling frequency of a video (shot detection results), we proceed by formulating the DYNAMIC BRIGHT-NESS SCALING PROBLEM (DBSP) in order to optimize the brightness while meeting the user acceptability of scaling magnitude and scaling temporal consistency.

Definition 1 (DBSP). Suppose that an input video is segmented into N shots by Algorithm 1, each shot with a duration of t_i , and will be scaled at each shot boundary. Given a smartphone with the power model constants φ and γ , as well as the threshold of user acceptability for pleasant viewing A (0 < A < 1) and the Weber constant C, the problem is to determine the brightness level b_i ($0 < b_i \leq 1$) for all shots i ($1 \leq i \leq N$) such that the display energy of the entire video is minimized under the constraints of individual shot brightness and inter-shot brightness down-scaling.

Mathematically, DBSP can be written as,

$$\begin{array}{ll} \min_{\mathbf{(b)}} & \sum_{i=1}^{N} \varphi b_{i}^{\gamma} t_{i} \\ \text{s. t.} & \mathcal{A} = \frac{1}{1 + \exp^{-(\alpha + \beta E M D_{shot, i} + \lambda b_{i})}} \ge A \\ & b_{i-1} - b_{i} \le C b_{i-1} \\ & 0 < b_{i} \le 1 \end{array} \tag{8}$$

where the second constraint enforces that down-scaled variations $(b_{i-1} > b_i)$ should not be too abrupt while up-scaled variations $(b_{i-1} < b_i)$ has no specific constraint. Considering the objective and the second constraint, DBSP is a nonlinear and non-convex problem. To efficiently solve this problem, we identify an important property of DBSP.

- **Theorem 1.** DBSP is a geometric programming that can be converted into a convex optimization problem.
- **Proof.** A geometric programming (GP) problem can be defined as the following standard form [35].

$$\min_{(\mathbf{x})} \quad \begin{array}{l} f_0(\mathbf{x}) \\ \text{s. t.} \quad f_u(\mathbf{x}) \le 1, \ u = 1, 2, \cdots, m \\ g_v(\mathbf{x}) = 1, \ v = 1, 2, \cdots, p \end{array}$$
(9)

where $f_u(\mathbf{x}) = \sum_l d_{ul} x_1^{a_{ul,1}} x_2^{a_{ul,2}} \cdots x_n^{a_{ul,n}}$ are posynomials and the coefficients satisfy $d_{ul} > 0$ and $a_{ul,j} \in \mathbf{R}$ for $j = 1, 2, \cdots, n$. Besides, $g_v(\mathbf{x}) = d_v x_1^{a_{v,1}} x_2^{a_{v,2}} \cdots x_n^{a_{v,n}}$ are monomials and the coefficients satisfy $d_v > 0$ and $a_{v,j} \in \mathbf{R}$ for $j = 1, 2, \cdots, n$.

In DBSP in (8), the objective is a posynomial since $\varphi t_i > 0$. The first constraint can be transformed into $\frac{\ln \frac{A}{1-A} - \alpha - \beta E M D_{shot,i}}{\lambda} b^{-1} \leq 1$. As $\alpha < 0, \beta < 0, \lambda > 0$ (according to Table 1) and $\ln \frac{A}{1-A} > 0$, the first constraint is a monomial and also a posynomial. The second constraint can also be transformed into a constraint that is a posynomial. The problem in (8) can then be converted into the following optimization problem:

$$\begin{array}{ll} \min_{\mathbf{(b)}} & \sum_{i=1}^{N} \varphi b_{i}^{\gamma} t_{i} \\ \text{s. t.} & \frac{\ln \frac{A}{1-A} - \alpha - \beta E M D_{shot,i}}{(1-C) b_{i}^{-1} b_{i-1} \leq 1} \\ & 0 < b_{i} \leq 1 \end{array} \tag{10}$$

which is a GP problem. Thus we prove DBSP is a GP. \Box

4.2 The Proposed Optimal Algorithm

In order to solve the GP in (10), one can transform the GP that is a non-convex problem into a convex optimization problem by replacing variables $b_i = e^{B_i}$ and taking the natural logarithms of both the objective function and the constraints [35]. The resulted convex problem can be routinely solved by a standard optimization solver. The best time complexity achieved by interior-point methods for a GP in (9) can be given by $\mathcal{O}((l+m)^{1/2}(ml^2 + l^3 + n^3))$ [36].

In this paper, however, we propose a new optimal algorithm that is specially designed for the GP in (10) to obtain the solution much more efficiently. The proposed algorithm will output the optimal brightness levels to minimize the display energy in CrowdDBS. The reason for developing this algorithm is that it is always desired to improve the time complexity of optimal algorithms and enhance system efficiency. Moreover, standard solvers are not always available in all software platforms in practice.

The proposed algorithm is summarized in Algorithm 2. Initially, the brightness of each shot b_i is set as the minimum acceptable level based on the first constraint in (10). Then, the brightness assignment iterates N times, where N is a constant predefined value representing the number of shots in the given video. Each time we pick the shot i_{max} that has the maximum current brightness level $b_{i_{max}}$ and assign its current brightness as its optimal brightness level $b_{i_{max}}$, opt. This optimal brightness level will not be changed and the shot i_{max} will be marked as visited. At the same time, we examine the next shot of shot i_{max} , i.e., shot $i_{max} + 1$. We

Algorithm 2 Optimal DBS Algorithm

1: **Initialize** N as a constant value \triangleright The number of shots 2: $\forall i, b_i \leftarrow \frac{\ln \frac{\Lambda}{1-A} - \alpha - \beta EMD_{shot,i}}{\lambda}$ 3: while $\exists b_i \neq 0$ do $\triangleright N$ iterations $i_{\max} \leftarrow \arg\max_i \{b_i | i = 1, 2, \cdots N\}$ 4: 5: $b_{i_{\max},opt} \leftarrow b_{i_{\max}}$ $b_{i_{\max}} \leftarrow 0$ ▷ Mark this shot as visited 6: 7: if $b_{i_{\max}+1} \neq 0$ then 8: $b_{i_{\max}+1} \leftarrow \max(b_{i_{\max}+1}, (1-C)b_{i_{\max},opt})$ ٩. end if 10: end while 11: return bopt

then update its brightness in order to ensure that the intershot brightness down-scaling is smooth based the second constraint of (10). Specifically, since $b_{i_{\max}} > b_{i_{\max}+1}$ before the update, the brightness of shot $i_{\max} + 1$ can be decreased at most to $(1-C)b_{i_{\max},opt}$ and should also be kept above its current value. Note that there is no need to examine the previous shot $i_{\max} - 1$ because $b_{i_{\max}-1} - b_{i_{\max}} \leq 0 \leq C b_{i_{\max}-1}$, which automatically satisfies the second constraint of (10). The algorithm always stops in N iterations and generates an optimal output after all shots are visited.

Theorem 2. Algorithm 2 is optimal for DBSP and can minimize the display energy.

Proof. The initialization of Algorithm 2 guarantees that the brightness of each shot is lower-bounded by the smallest acceptable scaling magnitude. Given that we visit and finalize shot i_{\max} that has the highest current brightness at every iteration, the brightness update of shot $i_{\max+1}$ based on the variation constraint will not be conflicted in the future iterations. This is because $b_{i_{\max}-1} < b_{i_{\max}}$ always holds in the future iterations and thereby the brightness of shot $i_{\max+1}$ will never be required to increase again. Therefore, $\mathbf{b_{opt}}$ is a feasible solution.

We now prove the optimality of Algorithm 2 by contradiction. Suppose there exists a feasible solution b' such that $\sum_{i=1}^{N} \varphi b_i^{\gamma} t_i < \sum_{i=1}^{N} \varphi b_{i,opt}^{\gamma} t_i$. Since t_i is fixed by the shot detection for both solutions, there is at least one shot that satisfies $\varphi b_i^{\gamma} t_i < \varphi b_{i,opt}^{\gamma} t_i$, i.e., $b_i' < b_{i,opt}$. If we denote the first shot that meets $b_i' < b_{i,opt}$ during the iterations of Algorithm 2 as shot *i*, then at the time point when shot *i* is marked as visited, we have $b_i' < b_{i,opt}$ for shot *i* and $b_j' \ge b_{j,opt}$ for visited shots $\{j|b_j = 0\}$. We now discuss two possible cases at this iteration.

 If b_{i-1} ≠ 0, i.e., shot i − 1 has not been visited, we must have not updated the brightness of shot i. Accordingly, we have

$$b'_i < b_{i,opt} = b_i = \frac{\ln \frac{A}{1-A} - \alpha - \beta EMD_{shot,i}}{\lambda}$$

which contradicts the acceptability constraint of individual shot, i.e., the first constraint of (10).

2) If $b_{i-1} = 0$, i.e., shot i - 1 has been visited, we may have updated the brightness of shot *i*. Hence, the



Fig. 6. Modular architecture of CrowdDBS in Mobile Video Streaming.

brightness level should satisfy

$$b'_{i} < b_{i,opt}$$

$$= \max\left(\frac{\ln \frac{A}{1-A} - \alpha - \beta EMD_{shot,i}}{\lambda}, (1-C)b_{i-1,opt}\right)$$

$$\leq \max\left(\frac{\ln \frac{A}{1-A} - \alpha - \beta EMD_{shot,i}}{\lambda}, (1-C)b'_{i-1}\right)$$

which contradicts either the first or the second constraint of (10).

To summarize, b' is not a feasible solution in both cases, which contradicts the assumption. Therefore, we can prove that Algorithm 2 is optimal for DBSP. \Box

Considering the N iterations and the $\arg \max$ operation at each iteration, we can arrive at the following conclusion about the complexity of the algorithm.

Proposition 1. The proposed algorithm 2 can solve DBSP in $\mathcal{O}(N^2)$.

Therefore, CrowdDBS can efficiently obtain the optimal brightness for a given video.

5 CROWDDBS IN VIDEO VIEWING SYSTEMS

In this section, we demonstrate how to integrate CrowdDBS into video viewing systems. We present two example use cases of CrowdDBS in online video streaming and offline video viewing.

Mobile Video Streaming. We first introduce a cloudassisted video streaming architecture that can effectively integrate CrowdDBS and reduce mobile video display energy. To perform real-time DBS for a given video, the powerful video cloud servers will analyze the content and determine the scaling frequency and magnitude while satisfying the brightness variation requirement. As shown in Fig. 6, when a video is uploaded, the server first analyzes the video by seeking its shot boundaries via Algorithm 1. It then optimizes the brightness for each video shot and ensures the brightness variation by solving the DBS optimization of (10) via Algorithm 2. Finally, a DBS profile will be generated for the video. The cloud also encodes the videos by standard H.264/AVC codec such that a group of picture (GoP) is the minimum coding unit that can be individually decoded. The video file, along with its associated DBS profile, are stored as regular web file for HTTP-based video streaming.

On the client side, when a user selects a video for streaming, the client will first download the associated DBS profile of that video unless the profile file has already been cached in the local storage. The client will then start to load the video file via HTTP GET. Once all data for a GoP is downloaded, the client can decode the frames belonging to this GoP. The decoded frames are then passed to a playback buffer and they can be played out only when the buffer is large enough to ensure the smooth playback. Note that the playback can start without completely downloading the entire video. During the video playback, dynamic brightness scaling will be carried out and synchronized with the playback progress based on the DBS profile.

Unlike on-board DBS analysis where repeated computation is needed for every device that requests the same video, such an architecture mitigates the DBS overhead on massive number of devices into a one-time cloud processing. Therefore, CrowdDBS-enabled mobile video streaming shall minimize the network-wide mobile display energy from the system point of view.

Offline Video Viewing. CrowdDBS can also be embedded into offline video viewing using the standalone mobile device. The key difference from online video streaming is that the DBS computation is completed by the local video client. Before a video can play with DBS service for the first time, the video client on the mobile device will conduct the same DBS computation as in the cloud servers and generate a DBS profile file for this video. Depending on users' tolerance of initial waiting, the DBS computation time may impact the viewing experience of some users. In practice, a possible solution is to limit the initial waiting time by a user option and to quickly start playing the video without a DBS service. At the same time, the DBS computation will be pushed into the background and the DBS service can be activated when the computation is completed. Note that the DBS computation is only needed once for one particular video, but the energy saving will be achieved every time the video is viewed in the future. Regarding the energy overhead of CrowdDBS, we will show in Section 6 that offline video viewing using CrowdDBS can still achieve a promising percentage of energy reduction.

By adopting the above architecture, CrowdDBS can serve as an advanced energy-saving service in modern online/offline mobile video systems. Note that CrowdDBS is also backward compatible in a sense that if the energysaving service is not required, the framework can be reduced to conventional mobile video streaming/viewing. Upon user input, CrowdDBS client simply applies a special DBS profile with brightness always equal to 1.0.

6 EVALUATIONS

In this section, we start with validating the proposed shot detection algorithm and the perception-brightness model. We then present extensive experiments evaluating the endto-end system under various practical situations. All experiments are carried out in real-world environment.

6.1 Experimental Setup

We have implemented CrowdDBS into both an HTTP progressive streaming system and an offline video viewing system. For the streaming system, we do not choose adaptive bitrate streaming because we intend to isolate other variables than brightness and keep the encoding quality fixed. This allows a more accurate study of user experience for brightness scaling. The video server is built on a university server via Apache 2, which manages the web access and generates the DBS profiles. To guarantee that the user experience is only impacted by DBS, we focus on a high-speed WiFi scenario. Before every streaming session, we confirm the bandwidth is larger than video bitrate (\sim 2Mbps) via *iPerf*. Thereby, video re-buffer and the potential impact of transmission can be excluded. For the offline system, we embed CrowdDBS into the smartphone's video player and thus enable DBS for viewing local videos. We prepare 15 clips (average length 126 seconds) using the same encoding setting in the three crowdsourcing studies. These videos can be categorized into 5 groups based on the average EMD_{shot} of the video.

The performance metrics of the evaluations include: 1) user acceptability. We evaluate the user acceptability of a given video via user studies with 20 participants. Since full videos with multiple shots are evaluated, we follow the test protocol as in the scaling frequency study. 2) energy reduction. We measure the system power consumption for accessing and displaying a video via Qualcomm Trepn Profiler, which reads the data from mobile hardware directly. It has been reported that Trepn Profiler can deliver an extremely close result (99% accuracy) as traditional power meter [37]. Our devices, including LG Optimus G Pro (default) and Google Nexus 6, are recommended by Qualcomm since accurate power readings can be measured on them [38]. During the measurement, all the unused Apps and services are closed. We also turn off unnecessary networking activities, e.g., automatic update and Bluetooth, and only keep the video transmission over WiFi. By multiplying the power by the length of the viewing session, the energy consumption can be derived. Finally, we compare the energy of the DBS-enabled playback against the full-brightness non-DBS playback in order to obtain the energy reduction of a specific DBS scheme. The full-brightness playback represents the enforced setting of mainstream mobile video Apps. After running the measurement for 10 times, we report the average results.

We compare CrowdDBS with two traditional DBS schemes that compute the perception-irrelevant signal analysis. The first benchmark, denoted by Var-SSIM, is based on [27], where the scaling frequency is determined by the variance among each frames' average luminance and the brightness level of a video segment is derived by continuously decreasing the brightness (at a step of 0.05) until the SSIM of one frame hits the requirement (i.e., 0.9). Furthermore, we also evaluate a system, denoted by Con-PSNR, based on [14]. It utilizes a constant number of frames as a scaling segment and derives the minimum supported brightness by meeting the PSNR threshold of 35 dB. For CrowdDBS, the acceptability threshold A is set to be 0.9. The Weber constant for brightness variation is usually obtained by phycological studies and has been widely studied. We adopt a classic Weber constant C of 0.14 [34].

6.2 Validating the Shot Detection Algorithm

We now tune and validate the proposed shot detection algorithm in order to obtain the DBS frequency accurately. As broadly used in shot detection research [28], we experimentally derive the global threshold θ and the relative threshold η . By going through all source videos, one obvious trend



Fig. 7. Overall shot detection accuracy of all videos.

TABLE 2 Detection false hits/miss rate (%)

sniper	bunny	birdman	boyhood	Budapest
0/8.06	4.72/2.36	2.63/18.42	0/2.86	0/4.90
imitation	selma	tear	theory	whiplash
5.00/7.00	2.07/2.07	11.03/5.15	6.06/0	1.57/3.15
ducks	oldtown	parkrun	rushhour	redkayak
0/0	0/0	0/0	0/0	0/0

we observed is that the EMD of true hard-cut positions is at least 13.37 times greater than that of their prior neighbors. Nonetheless, the EMD at non-hard-cut positions is at most 3.21 times greater. Hence, η is conservatively set as 10, to account for potential hard cuts with relatively small η . We then evaluate the algorithm under various θ . We show the overall false hits rate and miss rate of all video sources in Fig. 7. It can be seen that the false hit decreases as θ increases. This is because less EMD values can satisfy the global threshold, which reduces false hits. However, the chance of missing a true shot is simultaneously increased. As there is a stable trend for both curves, the optimal θ can be safely determined as 4.

By adopting the above optimal thresholds, we evaluate all the videos and show the detection accuracy in Table 2. In most cases, the results satisfy the typical requirement of 5% false hits and miss rate [28]. More importantly, we emphasize that false hits are usually found during a special lighting in the video, e.g., many camera lights shine repeatedly. Hence, a false hit is not quite dangerous as users can hardly perceive a brightness variation under those strong light effects. On the other hand, the luminance of missed shots is usually very similar to their neighbors, e.g., a man facing two shooting angles. Even though the missed shots were detected, we would assign them a brightness level similar to their neighbors. Therefore, we can conclude that the proposed detection algorithm can determine the DBS frequency satisfactorily.

6.3 Validating the Scaling Magnitude Model

In this section, we evaluate how well the predicted acceptability of the proposed model (5) fits the crowdsourced data.

We first perform likelihood ratio test of model fit between our model and the perfect model exploiting all degrees of freedom. The insignificant results ($\chi^2 = 454.1 df =$ 1963 p > 0.999) implies that there is very little unexplained variance and thus the proposed model fits well. To measure the similarity and correlation between predicted acceptability and the actual acceptability, we further calculate a set of metrics in Table 3, i.e., root-mean-square error, Nagelkerke pseudo R-squared, Pearson correlation coefficient (r) and Spearman rank correlation coefficient (rho). The rho

0.915

TABLE 3 Model Evaluations RMSE R^2

0.079

0.909

0.985



Fig. 8. Power breakdown of CrowdDBS-enabled video streaming system and offline viewing system.

crowdsourced model reaches a small RMSE and close-toone values for the other metrics, indicating an accurate and reasonable model. We also compute the correlation between ground truth and conventional objective DBS metrics, i.e., PSNR and SSIM between scaled and original images. The luminance of scaled images is computed by multiplying brightness level by original pixel values. Then the frameaveraged PSNR or SSIM is adopted for a shot. We observe that the crowdsourced model is 9%~23% more correlated to the user data. This manifests the large space to harness human perecption and save energy in CrowdDBS.

6.4 Overhead of CrowdDBS in Video Viewing Systems

In this section, we evaluate the overhead of CrowdDBS on mobile devices under both mobile video streaming and offline video viewing systems.

Mobile Video Streaming. The client-side overhead in the DBS-enabled cloud-assisted streaming architecture stems from the transmission of the DBS profile, which will incur additional transmission delay and energy consumption. To explore this issue, five users were asked to view the video "ducks" in a residential WiFi and a campus WiFi environment. They were asked to identify if they perceive significant start-up delay compared with the normal video streaming without DBS and if the initial delays under the two WiFi scenarios are different. We report that no noticeable delay was found in both cases. In fact, this is expected because the profile file is only about 1 KB. According to our measurement, the download time for the profile is at most a few hundreds of milliseconds, which can hardly be perceived during the initial video buffering. Regarding the extra energy, we will demonstrate later that such overhead is also marginal and will not undermine the performance gains of CrowdDBS.

Offline Video Viewing. To benchmark the on-device computation overhead of CrowdDBS, we compare the energy performance when CrowdDBS is applied into an offline viewing system and an online streaming system. By following the methodology of component-wise power measurement in [3], we first measure the power breakdown of CrowdDBS in offline video viewing. For the second system, we remove the on-device DBS computation of CrowdDBS and instead use the cloud-assisted architecture to carry out



Fig. 9. Energy reduction and user acceptability under different content luminance categories on G Pro.

the same analysis. We show the results of the two systems against the full-brightness non-DBS system in Fig. 8.

We observe that using CrowdDBS can significantly reduce the display energy in both online streaming and offline viewing systems. The cloud-assisted streaming system avoids the on-device DBS computation power for video analysis, which saves energy from an end-to-end perspective and reaches a 19% device energy reduction compared to the streaming system without CrowdDBS. On the other hand, due to the on-device video processing, the overhead of CrowdDBS is indeed increased in the offline video viewing system, diminishing the display energy saving. However, since the overall device energy of offline viewing also decreases (no WiFi transmission), CrowdDBS still arrives at a promising percentage of overall energy saving. In fact, the device energy reduction of CrowdDBS in offline viewing system is 17%, which is comparable to that in online streaming system. Moreover, such overhead is only introduced once for a given video when generating the DBS profile. Therefore, CrowdDBS can effectively minimize the display energy of both video streaming and offline viewing.

6.5 Impacts of Content on System Performance

We now show the results of CrowdDBS streaming under different practical settings.

The systematic performance of CrowdDBS when streaming different categories of content is shown in Fig. 9. We observe that CrowdDBS achieves 39% device energy reduction on average, or 39% and 56% more saving than Var-SSIM (28%) and Con-PSNR (25%), while still maintaining a satisfactory user experience. The energy saving is attributed to the crowdsourced model that effectively captures the lower bound of brightness from the perspective of human perception, which guides the energy minimization. Similarly, the shot-based scaling frequency and the DBS consistency constraint further guarantee the user experience.

In contrast, the benchmark systems using objective distortion metrics sacrifice much space in leveraging DBS and thus cannot maximally reduce the energy. Unfortunately, the relatively high brightness of benchmark systems does not improve the user experience. This is because there is virtually no perceptual difference between the brightness applied by CrowdDBS and the benchmark systems, making the high brightness unnecessary. Furthermore, users in benchmark systems constantly suffer from the poor alignment of scaling points and the lack of scaling consistency.

It is also interesting to see that the energy saving of CrowdDBS increases as the content becomes darker (higher category index). This results from the content-dependent QoE-brightness model in (5). In fact, the lowest acceptable brightness of darker content is smaller and thus CrowdDBS







can generally apply a more aggressive brightness scaling. Hence, CrowdDBS can explore the features of different video contents and minimize their energy accordingly.

6.6 Impacts of Device on System Performance

In addition to the default evaluation using the LCD-based LG G Pro, we further evaluate the system performance on an OLED smartphone, Google Nexus 6. We illustrate the results in Fig. 10. It can be seen that CrowdDBS still performs satisfactorily (28%~42% energy reduction with satisfactory experience) and presents a similar performance trend as in the case of LCD phones. This is because the proposed DBS optimization framework is modeled and solved without any device-specific assumptions, which allows CrowdDBS to be broadly applied in a device-neutral way.

Indeed, the amount of energy saving is generally smaller on Nexus 6. This is due to the distinct hardware and kernel efficiency across the devices, which results in the different model constants in the power model. In fact, as long as there is a reasonable power difference at different brightness levels, which is always true for modern devices, we can achieve desirable performance gain.

6.7 Impacts of Minimum Shot Length

The length of video shots directly determines the DBS frequency, which in turn impacts both user experience and energy consumption. As shown in Fig. 11, we demonstrate the system performance for streaming the same video under different L_{min} , the minimum possible shot length in Algorithm 1. In general, the energy reduction mitigates and the user experience improves as the minimum shot length increases. This verifies our previous argument that dividing a video into less shots (greater L_{min}) will render a more stable playback whereas the potential of manipulating DBS dynamics to save energy becomes smaller. Hence, by adjusting this important parameter, users can strike a customized tradeoff between user experience and energy. This provides a flexible user knob for different application scenarios.

7 DISCUSSION

Luminance Compensation. To compensate the brightness scaling, several works have strived to increase the video luminance before scaling the brightness [9]. We choose not to adopt such luminance compensation since our goal is to investigate the intrinsic relationship among user experience, brightness level, and content luminance at the fundamental level. Employing additional enhancement approaches, e.g., luminance compensation that builds on objective or linear assumptions, would obstruct this objective. In fact, part of video pixels can be fully compensated after luminance compensation whereas how the resulted distortion in the brightness-scaled video would affect human perception is still unknown. It is true that luminance compensation is a complementary approach to brightness scaling. Nevertheless, a separate full-scale research is indeed required. If the correlation between user experience and luminance compensation can be fully understood, even more energy savings can be achieved.

Viewing System Design. In the video streaming system implementation with CrowdDBS, an additional profile file is sent by the server to instruct proper DBS at the device. Alternatively, we can embed the profile information into the encoded video bitstream as a supplemental enhancement information of H.264 [12], [13], which can serve as an elegant way to communicate this small yet critical information. Note that the proposed server-driven metadata-assisted architecture has been actively studied by MPEG to reduce encoding, decoding and display energy of mobile streaming [39].

Bitrate adaptation and brightness scaling are two different dimensions of user experience. For example, if a user views a bitrate-adaptive video with CrowdDBS, the overall experience would be unpredictable since the CrowdDBS is obtained under one specific bitrate. Although adaptive bitrate streaming is not currently implemented in this research, we can achieve this easily by repeating the same user study to acquire the perception models under different bitrates. Similarly, DBS strategies under unusual mobile viewing environments, such as brighter outdoor or darker interior, can also be obtained.

Incentives for Improving CrowdDBS. After CrowdDBS is deployed, a larger scope of users can use CrowdDBS and report their experience in using CrowdDBS. Such new user data will be used to fine-tune CrowdDBS models in order to improve the system in practice. Services as an incentive [40] can be used toward future efforts in improving CrowdDBS. That is, if a user wants to benefit from the DBS service provided by CrowdDBS to save display energy, she also has to contribute to the system by reporting her experience using the current version of CrowdDBS. An incentive mechanism can be designed to ensure the fairness of multiple users in consuming the service while guaranteeing the quality of information provided by the users.

8 CONCLUSION

In this paper, we have introduced a new human-in-theloop approach for display energy reduction in mobile video viewing systems from the perspective of crowdsourced human perception modeling. Inspired by the insightful results of user experience under dynamic brightness scaling, we propose a suite of designs to handle brightness scaling frequency, magnitude and temporal consistency. Through optimizing the DBS service and integrating CrowdDBS into real-world video viewing systems, we show that the CrowdDBS framework can save 37% device energy on average while guaranteeing a satisfactory user experience.

More importantly, this research represents a paradigm shift in designing DBS schemes to overcome the prevalent issue related to display energy reduction at the user and application levels. We believe CrowdDBS shall open a new avenue of research to a suite of future interdisciplinary research strategies that integrate the human perception principles with different technologies from mobile computing, wireless communications, and multimedia processing.

9 ACKNOWLEDGEMENT

This research is supported by NSF ECCS-1405594 and ECCS-1406154 as well as National Science Foundation of China No. 61601080 and 61761136019.

REFERENCES

- Cisco, "Cisco visual networking index, 2014-2019," http://goo.gl/ v8pMqV.
- [2] X. Chen, Y. Chen, Z. Ma, and F. C. Fernandes, "How is energy consumed in smartphone display applications?" in ACM Workshop on Mobile Computing Systems and Applications, Feb. 2013.
- [3] M. Dong and L. Zhong, "Chameleon: A color-adaptive web browser for mobile OLED displays," *IEEE Trans. Mobile Comput.*, vol. 11, pp. 724–738, May 2012.
- [4] 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, pp. 25–32, Feb. 2004.
- [5] A. Iranli, H. Fatemi, and M. Pedram, "HEBS: histogram equalization for backlight scaling," in DATE, Mar. 2005.
- [6] A. Iranli, W. Lee, and M. Pedram, "HVS-aware dynamic backlight scaling in TFT-LCDs," *IEEE Trans. VLSI Syst.*, vol. 14, pp. 1103– 1116, Oct. 2006.
- [7] A. Bartolini, M. Ruggiero, and L. Benini, "HVS-DBS: human visual system-aware dynamic luminance backlight scaling for video streaming applications," in ACM EMSOFT, Oct. 2009.
- [8] 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, pp. 335–348, Feb. 2014.
- [9] Y. Liu, M. Xiao, M. Zhang, X. Li, M. Dong, Z. Ma, Z. Li, and S. Chen, "GoCAD: GPU-assisted online content-adaptive display power saving for mobile devices in internet streaming," in *International World Wide Web Conference (WWW)*, Apr. 2016.
- [10] Z. Yan and C. W. Chen, "RnB: rate and brightness adaptation for rate-distortion-energy tradeoff in HTTP adaptive streaming over mobile devices," in ACM International Conference on Mobile Computing and Networking (MobiCom), Oct. 2016.
- [11] Z. Yan, Q. Liu, T. Zhang, and C. W. Chen, "Exploring QoE for power efficiency: A field study on mobile videos with LCD displays," in ACM International Conference on Multimedia, Oct. 2015.
- [12] Q. Liu, Z. Yan, and C. W. Chen, "Cloud-based video streaming with systematic mobile display energy saving: Rate-distortiondisplay energy profiling," in *IEEE International Conference on Image Processing (ICIP)*, Sep. 2016.
- [13] Q. Liu, Z. Yan, and C. W. Chen, "Integrating mobile display energy saving into cloud-based video streaming via rate-distortiondisplay energy profiling," *IEEE Trans. on Cloud Comput.*, vol. 99, pp. 1–1, Nov. 2016.
- [14] L. Cheng, S. Mohapatra, M. El Zarki, N. Dutt, and N. Venkatasubramanian, "Quality-based backlight optimization for video playback on handheld devices," *Advances in Multimedia*, vol. 2007, no. 1, pp. 4–4, 2007.
- [15] P. Charonyktakis, M. Plakia, I. Tsamardinos, and M. Papadopouli, "On user-centric modular QoE prediction for VoIP based on machine-learning algorithms," *IEEE Trans. Mobile Comput.*, vol. 15, pp. 1443–1456, Jun. 2016.
- [16] J. Kim, P. T. Nguyen, S. Weir, P. J. Guo, R. C. Miller, and K. Z. Gajos, "Crowdsourcing step-by-step information extraction to enhance existing how-to videos," in ACM CHI, Apr. 2014.
- [17] M. Hoque, M. Siekkinen, and J. Nurminen, "Using crowd-sourced viewing statistics to save energy in wireless video streaming," in *ACM MobiCom*, Sep. 2013.

- [18] W. Song and D. Tjondronegoro, "Acceptability-based QoE models for mobile video," *IEEE Trans. Multimedia*, vol. 16, pp. 738–750, Apr. 2014.
- [19] S. Tavakoli, J. Gutierrez, and N. Garcia, "Subjective quality study of adaptive streaming of monoscopic and stereoscopic video," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 684–692, Apr. 2014.
- [20] K. Mitra, A. Zaslavsky, and C. Alund, "Context-aware QoE modelling, measurement, and prediction in mobile computing systems," *IEEE Trans. Mobile Comput.*, vol. 14, pp. 920–936, May 2015.
- [21] W. Song, D. Tjondronegoro, and M. Docherty, "Saving bitrate vs. pleasing users: where is the break-even point in mobile video quality," in ACM MM, Nov. 2011.
- [22] H. Knoche, J. D. McCarthy, and M. A. Sasse, "Can small be beautiful?: assessing image resolution requirements for mobile tv," in ACM MM, Nov. 2005.
- [23] G. Fechner, "Element of psychophysics," New York: Holt, Rinehart and Winston (Original work published 1860), 1966.
- [24] I. subjective video quality assessment methods for multimedia applications, "ITU-T Recommendation P.910," 1999.
- [25] DisplayMate, "Smartphone displays under bright ambient lighting shoot-out," http://goo.gl/n89Kgo.
- [26] W. Song, D. Tjondronegoro, and I. Himawan, "Acceptability-based qoe management for user-centric mobile video delivery: A field study evaluation," in ACM MM, Nov. 2014.
- [27] 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, pp. 398–405, Sep. 2004.
- [28] R. W. Lienhart, "Comparison of automatic shot boundary detection algorithms," in SPIE Electronic Imaging, 1998, pp. 290–301.
- [29] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, pp. 99–121, Nov. 2000.
- [30] A. M. Ferman, A. M. Tekalp, and R. Mehrotra, "Robust color histogram descriptors for video segment retrieval and identification," *IEEE Trans. Image Process.*, vol. 11, pp. 497–508, May 2002.
- [31] I. opinion model for video-telephony applications, "ITU-T Recommendation G.1070," 2007.
- [32] R. Christensen, Log-Linear Models and Logistic Regression. New York, USA: Springer-Verlag, 1997.
- [33] M. Zink, O. Künzel, J. Schmitt, and R. Steinmetz, "Subjective impression of variations in layer encoded videos," in *IEEE/ACM International Symposium on Quality of Service (IWQoS)*, Jun. 2003.
- [34] T. N. Cornsweet and H. Pinsker, "Luminance discrimination of brief flashes under various conditions of adaptation," *The Journal* of *Physiology*, vol. 176, pp. 294–310, Mar. 1965.
- [35] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, USA: Cambridge University Press, 2004.
- [36] A. Nemirovski, "Interior point polynomial time methods in convex programming," http://www2.isye.gatech.edu/~nemirovs/ Lect IPM.pdf.
- [37] M. A. Hoque, M. Siekkinen, K. N. Khan, Y. Xiao, and S. Tarkoma, "Modeling, profiling, and debugging the energy consumption of mobile devices," ACM Computing Surveys, vol. 48, p. 39, Feb. 2016.
- [38] Qualcomm, "Devices that report accurate battery power," http://goo.gl/ZecX96.
- [39] 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, pp. 80–87, Mar. 2015.
- [40] D. Yang, G. Xue, X. Fang, and J. Tan, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *ACM MobiCom*, Aug. 2012.



Zhisheng Yan received B.Eng. degree from Shandong University, China in 2010, M.Eng. degree from University of Science and Technology of China in 2013, and Ph.D. degree in Computer Science and Engineering from State University of New York at Buffalo in 2017. Currently, he is an Assistant Professor in the Department of Computer Science at Georgia State University. His research interests lie in mobile and networked systems, including content delivery systems, mobile health, and energy-saving mobile display.



Qian Liu received her B.S. and M.S. degrees from Dalian University of Technology, Dalian, China, in 2006 and 2009, respectively, and the Ph.D. degree from The State University of New York at Buffalo (SUNY Buffalo). She was a Postdoctoral Fellow at the Ubiquitous Multimedia Laboratory, SUNY Buffalo from 2013 to 2015. She received the Alexander von Humboldt Fellowship in 2015, and is now a Postdoctoral Fellow at the Chair of Media Technology and the Chair of Communication Networks Technical University

of Munich, Germany. She is also a Research Associate Professor at the Dept. of Computer Science and Technology, Dalian University of Technology, China.

Her current research interests include multimedia transmission over MIMO systems, IEEE 802.11 wireless networks and LTE networks, device-to-device communication, energy-aware multimedia delivery, and the Tactile Internet. She received the Best Paper Runner-up Award at the 2012 International Conference on Complex Medical Engineering and was in the finalist for the Best Student Paper Award at the 2011 IEEE International Symposium on Circuits and Systems.



Tong Zhang (M'02;SM'08) received the BS and MS degrees in electrical engineering from Xian Jiaotong University, China, in 1995 and 1998, respectively. He received the PhD degree in electrical engineering from the University of Minnesota, Minneapolis, in 2002. Currently he is a professor with Electrical, Computer and Systems Engineering Department, Rensselaer Polytechnic Institute, Troy, New York. His current research interests include circuits and systems for data storage, signal processing, and computing.

data storage, signal processing, and computing. Currently, he serves as an associate editor for the IEEE Transactions on Circuits and Systems-II and the IEEE Trans. on Signal Processing.



Chang Wen Chen (F'04) received his BS from University of Science and Technology of China in 1983, MSEE from University of Southern California in 1986, and Ph.D. from University of Illinois at Urbana-Champaign in 1992. He is currently an Empire Innovation Professor of Computer Science and Engineering at the University at Buffalo, State University of New York. He was Allen Henry Endow Chair Professor at the Florida Institute of Technology from July 2003 to December 2007. He was on the faculty of

Electrical and Computer Engineering at the University of Rochester from 1992 to 1996 and on the faculty of Electrical and Computer Engineering at the University of Missouri-Columbia from 1996 to 2003.

He has been the Editor-in-Chief for IEEE Trans. Multimedia since January 2014. He has also served as the Editor-in-Chief for IEEE Trans. Circuits and Systems for Video Technology from 2006 to 2009. He has been an Editor for several other major IEEE Transactions and Journals, including the Proceedings of IEEE, IEEE Journal of Selected Areas in Communications, and IEEE Journal on Emerging and Selected Topics in Circuits and Systems. He has served as Conference Chair for several major IEEE, ACM and SPIE conferences related to multimedia, video communications and signal processing. His research is supported by NSF, DARPA, Air Force, NASA, Whitaker Foundation, Microsoft, Intel, Kodak, Huawei, and Technicolor.

He and his students have received eight (8) Best Paper Awards or Best Student Paper Awards over the past two decades. He has also received several research and professional achievement awards, including the Sigma Xi Excellence in Graduate Research Mentoring Award in 2003, Alexander von Humboldt Research Award in 2010, the University at Buffalo Exceptional Scholar - Sustained Achievement Award for 2012, and the State University of New York Chancellors Award for Excellence in Scholarship and Creative Activities in 2016. He is an IEEE Fellow and an SPIE Fellow.