

# Exploring QoE for Power Efficiency: A Field Study on Mobile Videos with LCD Displays

Zhisheng Yan, Qian Liu  
State University of New York  
at Buffalo

Tong Zhang  
Rensselaer Polytechnic  
Institute

Chang Wen Chen  
State University of New York  
at Buffalo

## ABSTRACT

Display power consumption has become a major concern for both mobile users and design engineers, especially considering the prevalence of today's video-rich mobile services. The power consumption of liquid crystal display (LCD), a dominant mobile display technology, can be reduced by dynamic backlight scaling (DBS). However, such dynamic changes of screen brightness may degrade users' quality of experience (QoE) in viewing videos. How would QoE be impacted by different DBS strategies has not yet been understood clearly and thus obscures the way to achieve systematic power saving. In this paper, we take a first step to explore the QoE of DBS on smartphones and aim at maximally enhancing the display power performance without negatively impacting users' QoE. In particular, we conduct three motivational studies to uncover the inherent relationship between QoE and backlight scaling frequency, magnitude, and temporal consistency, respectively. Motivated by the findings of these studies, we design a suite of techniques to implement a comprehensive DBS strategy. We demonstrate an example application of the proposed DBS designs in a mobile video streaming system. Measurements and user evaluations show that more than 40% system power reduction, or equivalently, 20% more power savings than the non-QoE approaches, can be achieved without QoE impairment.

## Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Video

## General Terms

Design, Experimentation, Human Factors

## Keywords

LCD, Mobile Video Streaming, Power Consumption, QoE

## 1. INTRODUCTION

The proliferation of mobile web access and social networks has dramatically boosted mobile video services. Although

mobile computing and storage capabilities have been continuously expanding in order to meet the ever-growing video services, relatively short lifetime of batteries has long been a major complaint by smartphone users. Among the operational components of a mobile phone, display power is one crucial issue that needs to be addressed urgently. Not only have displays been identified as one of the most power-consuming subsystems [1], but display power is also consumed in a broader video-related applications as long as the video is playing on the mobile phone. For example, unlike communication components that consume energy only when video streaming is in session, power dissipation of displays shall occur in both online and offline video viewing.

Liquid crystal display (LCD) is a dominant display technology that is adopted by larger number of modern devices, e.g., Apple iPhone 6/6 Plus, Google Nexus 4/5, and LG G 3/4/Pro. According to a recent study [2], the revenue share of LCD-based mobile phones is expected to reach 70% in 2020. The power consumption of LCD primarily comes from the backlight at the display panel. By dimming the backlight, we could save the display energy. However, dimmed backlight may lead to a lower perceptual luminance of displayed pixels, which would cause video distortion and eventually degrade users' *quality of experience* (QoE). Hence, the key principle of LCD power reduction in mobile videos is to dynamically scale the backlight as low as possible without negatively impacting users' QoE. This is commonly termed as *dynamic backlight scaling* (DBS).

Despite this general consensus, we observe that our understanding of power-saving strategies for LCD is still limited. This might surprise many experts since DBS has been actively studied in design automation communities. The reason behind this observation is that mobile videos introduce new effects on the QoE of backlight scaling while little is known regarding the optimal backlight under this new context. First, prior backlight scaling schemes for a single image [3–5] cannot be directly applied to the multi-frame videos. Second, most existing works [6, 7] were implemented using programmable LCD modules, which cannot adequately reflect human perception on smartphones that employ larger screen size and higher pixel-density. Most importantly, prior objective distortion metrics for DBS, e.g., linear relationship between perceptual luminance and backlight levels [4, 8, 9], become inaccurate and need to be replaced by subjective QoE modeling that specifically targets DBS.

On the other hand, although video QoE has been extensively researched in the context of encoding and streaming to improve compression and delivery performance, QoE of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

MM '15 Brisbane, Australia

Copyright 2015 ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

backlight-scaled videos for power efficiency has not been studied. In this research, we take a first step to explore QoE when DBS is applied in order to maximally enhance the power performance of LCD-based mobile phones without sacrificing end users' QoE. To achieve this ambitious objective, we are facing several research challenges.

- Finding a suitable *scaling frequency*: How frequently should DBS be applied is fundamentally dependent on both QoE and power. Less frequent scaling will result in more stable viewing experience, but will leave much smaller space for manipulating the backlight dynamics, which diminishes the potential for power saving [8]. When to apply the scaling also matters because abrupt variation of backlight is annoying to users.
- Determining a smallest *scaling magnitude*: For a given segment of video frames, obtaining a smallest scaling magnitude in order to maximally reduce power consumption is a non-trivial task. The inherent relationship between QoE and backlight change has not yet been understood clearly. This is further complicated by the fact that users may express distinct perception for different video contents even under the same backlight condition.
- Applying *temporally consistent* DBS across multiple segments: Simply scaling each segment's backlight individually without inter-segment consideration shall lead to QoE degradation because human eyes are shown to be extremely sensitive to temporally inconsistent videos. For example, adaptive streaming studies [10] have reported that quality switches with high variation frustrate the end users.

In order to tackle these challenges, we have carried out a comprehensive investigation on the QoE of dynamic backlight scaling. We have developed a customized video App, QoEPlayer, to conduct three motivational studies to explore the connection between QoE and scaling frequency, magnitude, and temporal consistency, respectively. By leveraging statistical analysis, we discover that a video shot based scaling frequency is on average 11 times more acceptable by users than other alternative approaches. Accordingly, we design and tune a video shot detection algorithm to decide the best scaling frequency. Furthermore, we observe a content-luminance-dependent logistic relationship (instead of linear) between QoE and backlight levels. We then construct a QoE model using logistic regression in order to determine the best scaling magnitude of a given video shot. Moreover, we find it interesting that users are only sensitive to down-scaled backlight switches rather than up-scaled ones. Therefore, we train a binary classifier to detect those inconsistent down-scaled backlight switches and then properly smooth these variations. Finally, combining all these findings, we develop a systematic strategy for DBS as applied to mobile phones. We demonstrate an example application of the proposed designs in a video delivery system and show that more than 40% system power can be saved, which is equivalent to 20% more power savings than conventional strategies without QoE considerations.

To summarize, the contributions of the proposed research include:

- A open-source video App, *QoEPlayer*, for subjective mobile video studies on Android platform (Section 3).

- A set of motivational studies to explore the intrinsic relation between QoE and scaling frequency, magnitude and temporal consistency (Section 4-6).
- A suite of designs to achieve the most power-efficient DBS strategies without negative effects on user experience (Section 4-6).
- A practical demonstration of the usage of the proposed DBS designs, showing 40% system power reduction, or equivalently, 20% more power savings than non-QoE approaches (Section 7).

## 2. RELATED WORK

### 2.1 Dynamic Backlight Scaling

In general, exiting DBS schemes first compute the perceptual luminance of a backlight-scaled image by assuming it is linearly proportional to scaling magnitude and/or original pixel value. Then the backlight keeps scaling down until certain distortion thresholds between the two images are met.

Particularly, Cheng and Pedram scaled the backlight while preserving the image contrast [3]. Chang *et al.* increased the original pixel values to compensate for the scaling distortion and then retained the brightness and contrast [4]. Iranli *et al.* compensated the original image by considering histogram equalization [5]. These single-image level schemes demonstrate the principle of backlight scaling, but they are not directly applicable to mobile videos.

DBS for videos were studied by considering video-related distortion metrics. Structural similarity (SSIM) [7] and peak signal-to-noise ratio (PSNR) [11] based measure were exploited to apply DBS. Visual sensitivity characterization in frequency domain was used to smooth the flicker effect of DBS [6]. Recently, Lin *et al.* optimized the scaling strategies by dynamic programming to minimize power consumption [8]. Liu *et al.* proposed to use GPU to perform luminance compensation and to reduce power effectively [9]. However, these works all rely on objective distortion formulations that are not specifically developed for DBS and therefore lack direct relationship with user experience.

In this paper, we explore the user experience of backlight-scaled videos through subjective user tests on modern smartphones. We aim at uncovering the inherent relations among QoE, backlight and video content, and utilizing them to improve the power performance of mobile phones.

### 2.2 QoE Studies

Many objective metrics for QoE focus on low-level video features, e.g., PSNR and SSIM [12]. However, whether or not these objective metrics match well with user experience is still controversial and is usually application-dependent [13]. More commonly, QoE is modeled as mean opinion scores obtained from subjective evaluation data. Many efforts were made for different applications, such as general networked video systems [14], adaptive HTTP streaming [15], and ITU video-telephony [16]. One potential issue of such score-based tests is that participants may be overburdened by the multiple levels of score and may struggle with making a decision [17].

Recently, binary assessment choice was suggested for subjective tests to evaluate the acceptability of videos [13,17,18]. The lowest acceptable video for pleasant viewing can be

found using Method of Limits in classical psychophysics [19], where users view a series of videos varied in successive discrete steps with descending or ascending order and submit their binary acceptability choice.

There were also important data-driven works that mine the large-scale network traffic [20] or user engagement data [21] to model QoE in Internet applications. However, this is not feasible for DBS on mobile phones as users need to personally hold the device and view the backlight-scaled video to experience the video display.

All these exiting works explore how QoE is impacted by encoding or transmission parameters, e.g. resolution, bit-rate and delay, in order to enhance the networking performance. In this paper, however, our objective is fundamentally different in that we probe into the new QoE space of backlight scaling in order to enhance the power performance.

### 3. USER STUDY SETUP

In this section, we introduce the general setup of user studies and system demo. Test-dependent setup will be presented in the subsequent sections.

**Test Devices.** LG Optimus G Pro and Google Nexus 4 are used in all tests except the study of temporal consistency where G Pro and Xiaomi Mi 2 are used. The screen size, resolution, pixel density of G Pro, Nexus 4, and Mi 2 are 5.5 inch/1080x1920/401 ppi, 4.7 inch/768x1280/318 ppi, 4.3 inch/720x1280/342 ppi, respectively. The diverse screen setting allows us to study the impacts for different devices.

**Participants.** We recruit a total of 50 participants from two universities. Their ages range from 20 to 39. There are 32 males and 18 females. All participants have a normal color vision. We also report that one participant works in the area of mobile display. However, we do not find any significant difference in that particular result. A set of 40 participants take part in the three motivational studies whereas a set of 20 participants evaluate the system application. For each test, participants are divided into two halves and each half uses a different test phone.

**Video Sources.** We use 15 training sources for motivational studies. There are 5 standard sequences, 2 open movies [22], and 8 movie trailers that are 2015 Oscar nominees for best picture. To validate the designs in system-level application, we also prepare another 4 test sources that are 2015 Oscar nominees for best actor and actress without overlapping with above sources. Such selection concurs with the important principle of user studies that real-world materials should be used [17]. Since we are interested in DBS, we encode all videos identically at a high quality with 1280x720 resolution, 23 quantization parameter, 24 frames/second, and H.264/AVC baseline profile. We also encode all audio materials at AAC format with 44kHz and >180kbps to guarantee that users can understand the content and enjoy the viewing experience. These encoded sources will be split into test clips for the studies in the subsequent sections.

**Assessment Objective.** Due to the simplicity and comfort of user decision in binary-choice QoE tests, we follow prior works [13,17,18] to adopt this promising method. Participants are asked to identify whether or not a test clip is acceptable. Since we intend to maximally reduce the power without negative effects on users' regular viewing experience, acceptable quality is informed as the quality with tiny/no adjustment of screen brightness that participants would enjoy and thus have a pleasant experience in daily life.

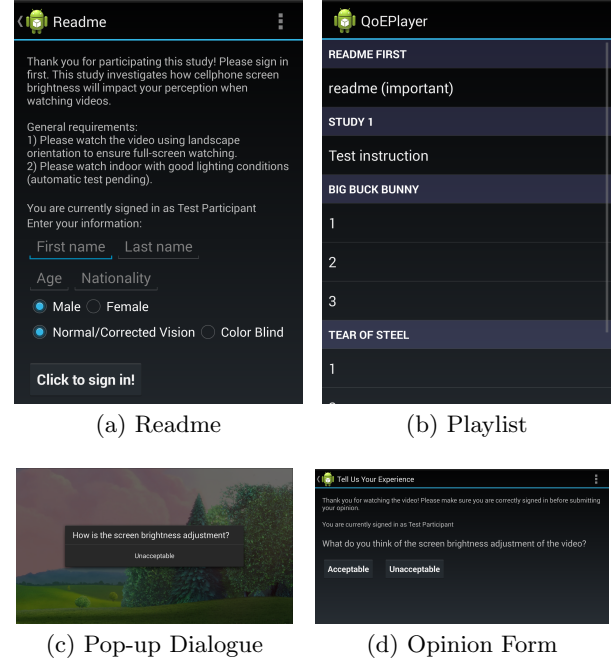


Figure 1: Screenshots of QoEPlayer.

**Test Environment.** Human perception of screen brightness might be affected by ambient light. We focus on the ambient light ranging from 100 to 1000 lux, which mimics the indoor residential and working lighting [23]. This is because people spend majority of time on viewing mobile videos in such indoor scenarios. According to the study in [24], 358 out of 663 people watch mobile video at home while 176 at university or working place, 92 at vehicle and 37 at cafe shop, barber and bus station. Although the location distribution within the categories is not known, it is still clear that a greater bulk of mobile viewing occurs indoor.

**Test App: QoEPlayer.** We implement an open-source<sup>1</sup> Android application for the tests. Initially, participants register at the Readme page (Figure 1a) by providing personal information. They are then asked to randomly select a video clip that has multiple versions differed by a certain stimulus (Figure 1b). The presentation order of multiple versions can be customized as random, ascending, and descending. Once a test version is selected, QoEPlayer will automatically check whether the ambient light meets the requirement via the light sensor in the device. If satisfied, video playback will start. Otherwise, a warning window that instructs users to move to a darker/brighter environment will pop up.

Upon playing the video, QoEPlayer creates a control thread to dynamically adjust the backlight via Android API based on a given scaling profile. It also guarantees that DBS is synchronized with the video and audio codec to support pause, fast forward, etc. During video playback, participants can double tap the screen to initiate a pop-up dialogue (Figure 1c) and submit their decision. Note that this function is only active when allowed by the test protocol. Alternatively, there is always an opinion form (Figure 1d) showing

<sup>1</sup>[http://www.cse.buffalo.edu/UBMM/People/zhisheng\\_yan.html](http://www.cse.buffalo.edu/UBMM/People/zhisheng_yan.html)

at the end of playback. After the choices are submitted in either way, QoEPlayer will automatically save the results of participants and return to the playlist.

This automatic process provides participants with a simpler and more immersive viewing experience, which eventually speeds up the studies. We emphasize that QoEPlayer can be modified and ported to any mobile QoE studies.

## 4. ADDRESSING DBS FREQUENCY

To find a proper scaling frequency, we must identify the requirements in terms of both QoE and power. First, it is desired to let the backlight scaling occur at time points when users pay much less attention to the video. Second, the number of video segments separated by the scaling points should be as many as possible if QoE is not compromised. In fact, more segments allow more rooms for manipulating the backlight dynamics, hence reducing more power [8]. For example, if we mix two segments whose lowest scaling magnitude to maintain QoE are  $\mathcal{M}_1$  and  $\mathcal{M}_2$ , we would need to apply the magnitude  $\max(\mathcal{M}_1, \mathcal{M}_2)$  to the mixed segment to ensure QoE, which is less power efficient than assigning  $\mathcal{M}_1$  and  $\mathcal{M}_2$  to each segment, respectively.

### 4.1 Motivational Study

#### 4.1.1 Methodology

Ideally, we should examine the QoE under all combinations of segmentation for a video and then model the scaling frequency versus QoE data and content features of segments. However, this is infeasible as there are too many ways of segmentation, e.g., a 1-second video will have tens of millions options. Therefore, we proceed by comparing several possible approaches and obtain a suboptimal yet acceptable solution. To the best of our knowledge, only two previous works considered DBS frequency. In [8], a constant scaling frequency with certain number of frames was used (denoted as *Const*). In [25], the authors used the rule that the variance of average luminance of frames belonging to a segment should be less than a threshold value (denoted as *GoS*).

In addition, we propose to adopt a *video shot* as one scaling segment and scale the backlight at shot boundaries. A shot refers to a series of consecutive frames played for an uninterrupted period of time. We propose this strategy as there is usually a sudden transition of content when the shot is switched. We anticipate that users might pay much less attention to the shot-boundary frames with sudden switch than to the within-shot frames showing the actual contents. Thereby, the impacts of DBS can be diminished. Note that video scenes in semantic level is a similar idea. We prefer shots as scaling segments since a scene consists of one or more shots and thus leaves us less space for power reduction.

#### 4.1.2 Results

We prepare 4 test clips (2 full trailers and 2 movie clips) with average duration 115 seconds. Each clip is dynamically backlight scaled using the above three frequencies. The constant frequency and variance threshold are set to the default value of 10 frames [8] and 40 [25], respectively. Besides, we use a shot detection algorithm to be introduced in the next subsection to find the shot boundaries and manually correct the false/missed detection. For all frequency strategies, the scaling magnitude  $\mathcal{M}$  of each segment  $i$  is decided using a simple linear mapping that is based on the scheme in [25].

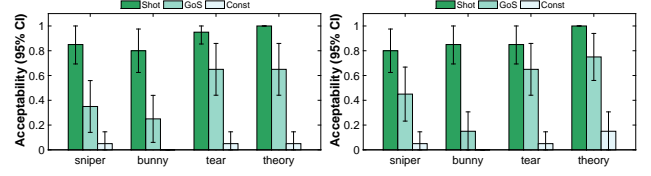


Figure 2: Acceptability under different scaling frequencies on G Pro (left) and Nexus 4 (right).

Specifically, if the content of the current segment is darker than the previous segment, its backlight magnitude can be a little less than the previous magnitude without degrading QoE and vice versa, i.e.,

$$\mathcal{M}_i = \begin{cases} \max(\mathcal{M}_{i-1} - \mathcal{M}_{step}, 0) & Y_i < Y_{i-1} \\ \mathcal{M}_{i-1} & Y_i = Y_{i-1} \\ \min(\mathcal{M}_{i-1} + \mathcal{M}_{step}, 1) & Y_i > Y_{i-1} \end{cases} \quad (1)$$

where  $Y$  is the average luminance of a segment over all frames and pixels, and  $\mathcal{M}_{step}$  is backlight variation step that is 0.2. We choose this  $\mathcal{M}_{step}$  to avoid too little or too much backlight variations that may make all or none frequencies strategies acceptable. Participants are asked to watch 3 frequency versions of a clip in a random order and then move to next clip. They can make a “unacceptable” decision during the backlight-scaled playback while they will need to wait until the end of playback to confirm “acceptable”.

To better analyze the binary subjective data, we use acceptability  $\mathcal{A}$  as in [13, 17, 24], i.e.,  $\mathcal{A} = \frac{N_{acc}}{N_{tot}}$ , where  $N_{acc}$  is the number of acceptable decisions and  $N_{tot}$  is the number of total votes. We show the acceptability (95% CI) of 4 clips under different scaling frequencies in Figure 2. We can observe that QoE is clearly impacted by frequency strategies. Chi-square test between frequency strategies and acceptability,  $\chi^2 = 167.008$   $df = 2$   $p < 0.001$ , further supports that QoE under different scaling frequencies differs significantly from each other in a statistical sense.

It can also be seen from Figure 2 that shot-based scaling frequency achieves the best result for all test clips on both devices. Its average acceptability reaches 88.75%, and is on average 1.82 times and 20.3 times better than the other two methods, respectively. This could be explained as users pay most attentions to the essential parts of a video, such as objects, people, and stories. Subconsciously, they do not quite care about the shot switches. Hence, frequent backlight scaling in the middle of presenting those essential contents is much more annoying than that in shot boundaries. Thereby, we can conclude that shot-based scaling frequency provides a simple yet effective solution to this challenging problem. It is also noted that QoE results are content-dependent, which shall guide us to deliberately set the scaling magnitude of shots based on the content characteristics.

### 4.2 A Shot Detection Algorithm for DBS

We now present the design of a new shot detection algorithm to determine DBS frequency. Video shot detection have been previously studied [26]. Although none of exiting designs focuses on DBS, they provide the foundations for the proposed algorithm. Shot detection usually consists of two steps [27]: computing discontinuity values for consecutive frames and detecting the shots based on the discontinuity.

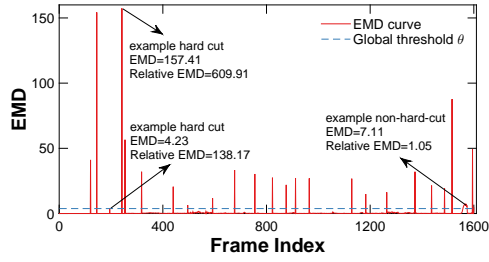


Figure 3: EMD of “sniper” (first 1600 frames)

**Algorithm 1** Proposed Shot Detection Algorithm

```

1: procedure SHOT DETECTION FOR DBS
2:    $S \leftarrow 0, EMD_{pre} \leftarrow EMD(0, 1)$ 
3:   for  $k \leftarrow 1; k < LastFrame; k \leftarrow k + 1$  do
4:     if  $EMD(k, k + 1) > \theta$  &  $\frac{EMD(k, k + 1)}{EMD_{pre}} > \eta$  then
5:       if  $k + 1 - s_{pre} \geq D_{min}$  then
6:          $S \leftarrow S \cup \{k + 1\}, s_{pre} \leftarrow k + 1$ 
7:        $EMD_{pre} \leftarrow EMD(k, k + 1)$ 
8:   return  $S$  ▷  $S$ : set of start frames for shots

```

We propose to adopt the luminance histogram based discontinuity metric in DBS. This is because DBS is more power efficient if we divide a video into segments with distinct luminance feature rather than mixing them together. Furthermore, luminance metric is robust to missed detection in the case of DBS since missed detection could occur only when the luminance strength between two shots are very similar. This is not a problem as we would assign the same scaling magnitude to both shots even if we detected them. Summarizing these discussions, we set the objective to detect hard-cut shots with abrupt luminance change. It is unnecessary to detect those shots with gradual transition like dissolves.

We employ Earth Mover’s Distance (EMD) [28], which is commonly used in image retrieval, as the luminance histogram feature. EMD is defined as the minimum cost paid to transform one histogram into the other. Formally, EMD between histogram  $P$  and  $Q$  is expressed as:

$$EMD(P, Q) = \frac{(\min_{f_{ij}} \sum_{i,j} f_{ij} d_{ij}) / (\sum_{i,j} f_{ij})}{s.t. \quad \begin{aligned} f_{ij} &\geq 0, \sum_j f_{ij} \leq P_i, \sum_i f_{ij} \leq Q_j \\ \sum_{i,j} f_{ij} &= \min(\sum_i P_i, \sum_j Q_j) \end{aligned}} \quad (2)$$

where  $f_{ij}$  is the flow amount transported from  $i$ th bin to  $j$ th bin and  $d_{ij}$  is ground distance between bin  $i$  and  $j$ . We choose EMD because it is a cross-bin metric that considers both bin height and inter-bin distance. Unlike bin-to-bin metrics used in conventional shot detection, EMD effectively captures the luminance discontinuity, which makes it a satisfactory metric for DBS.

The proposed shot detection exploits EMD for global thresholding and relative EMD for local thresholding to make the detection decision. This is more flexible than conventional global threshold based methods. The reason is that we observe a wide range of EMD values at true hard-cut positions, as exemplified in Figure 2. By incorporating relative threshold  $\eta$ , we can ensure that non-hard-cut positions with EMD greater than global threshold  $\theta$  would be dropped. Since most hardware cannot support per-frame scaling, the shot duration should also be greater than a minimum value  $D_{min}$ , which is decided to be 5 based on our measurement. Note

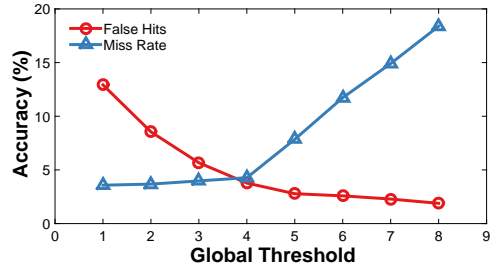


Figure 4: Overall shot detection accuracy

Table 1: Detection false hits/miss rate (%)

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

that shot detection is not sensitive to this hardware limit as most shots last more than 1 second and thus per-shot scaling can be easily supported. The algorithm is summarized in Algorithm 1. Specifically, we first calculate the EMD between frame  $k$  and its previous/next frame, respectively. If  $EMD(k, k + 1)$  is greater than global threshold and sufficiently larger than  $EMD(k - 1, k)$ , as well as current shot lasting more than  $D_{min}$ , we can decide that a new video shot starts from frame  $k + 1$ .

We find the proper thresholds by experiments, as commonly done in shot detection research [26, 27]. We observe that all hard-cut positions have a much larger EMD than their neighbors. We go through all 15 training sources and discover that EMD values at hard-cut positions are at least 13.37 times greater than its prior neighbor while the EMDs at non-hard-cut positions are at most 3.21 times greater. Hence, we conservatively set  $\eta$  to 10 for possible hard cuts that have relatively small relative EMD.

Next, we range the global threshold from 1 to 8 and show the overall detection accuracy of all video sources in Figure 4, wherein false hits is the ratio of falsely detected shots to the actual number of shots and miss rate is the ratio of missed shots to the actual number of shots. A larger  $\theta$  allows less EMD values to pass the global thresholding and thus is less likely for false hits. However, it would also increase the chance of missing true shots. Given that both curves show stable trends, we can safely determine the optimal  $\theta$  as 4. The detection results for each source under optimal thresholds are shown in Table 1. The accuracy fulfills the requirement of 5% false hits and miss rate [26] in most cases.

More importantly, we shall report that a false hit is usually found when there is a special lighting in the video, e.g., multiple camera lights repeatedly on. Thereby, false hits are not quite dangerous since people can hardly discern a backlight change under those strong lighting effects. On the other hand, the missed shots usually have very similar content luminance to its neighbors, e.g., a talking man from two shotting angles. Even though those shots are not missed, we would scale the backlight for them similarly as their neighbors. Based on the results and DBS’s insensitivity to shot detection accuracy, we conclude that the proposed algorithm can achieve a satisfactory performance for determining scaling frequency.



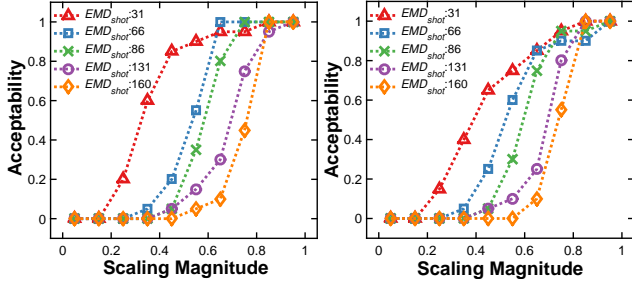


Figure 5: Acceptability under different scaling magnitude on G Pro (left) and Nexus 4 (right).

## 5. MODELING DBS MAGNITUDE

### 5.1 Motivational Study

We then proceed to determine the smallest possible scaling magnitude of a given video shot. Recall that both scaling magnitude and content luminance have strong effects on QoE. Hence, we must analyze the acceptability of shots with different content luminance under different backlight levels. It is desired to represent the luminance feature of a shot by using EMD because EMD has shown advantages in characterizing the luminance as discussed previously. However, we shall need to first obtain a shot’s histogram in order to compute EMD. Median histogram is a histogram descriptor for a group of frames and has been widely used in multimedia applications [29]. Each bin of the median histogram is computed by the median bin height over all the frame histograms. We select this descriptor since it can effectively eliminate the outlier frames within the shot [29]. Finally, we define the luminance feature of a shot, denoted by  $EMD_{shot}$ , as the EMD distance between the shot’s median histogram and the constant histogram of a black (zero-intensity) image. A larger  $EMD_{shot}$  indicates a brighter content.

We compute  $EMD_{shot}$  for all shots in all the training videos. A total of 1014 values are collected and they range from 16 to 185.38. For user study, we divide  $EMD_{shot}$  into 5 categories (category 5 darkest) with a step of 34 to represent different luminance strength of a shot. We then prepare 5 test clips accordingly (average duration 9 seconds). Besides, we vary the backlight scaling magnitude from 0.95 to 0.05 at a step of 0.1. The scaling magnitude can span from 0 (turn off) to 1 (full backlight). Overall, we prepare 50 versions of backlight scaled shots ( $5 EMD_{shot} \times 10 \mathcal{M}$ ) for this user study. Notice that the backlight is scaled only once at the beginning of the shot and the “acceptable” quality becomes the static brightness that users perceive no/tiny difference compared with the full-backlight version and would enjoy everyday. Participants are first asked to watch a full-backlight version as a reference. Then 10 scaled versions are played either from highest magnitude or from lowest magnitude using Method of Limit as in [13, 17]. Participants will choose whether or not each version is acceptable. They are allowed to confirm acceptable/unacceptable during the playback.

Eventually, we collect 2000 binary data points for this study (40 participants  $\times$  50 video versions). To better visualize the binary data and grasp a clear understanding of the data, we convert the binary data into acceptability as in the previous section. We show the raw acceptability data for 5

Table 2: Model Coefficients

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

luminance categories in Figure 5. It is clear that the relationship between QoE and backlight is not linear, indicating past assumptions, e.g., [4, 8, 9], is not accurate. Instead, we observe a sigmoid curve (“S” shape curve) for all luminance categories. Furthermore, the sigmoid curve is horizontally shifted when  $EMD_{shot}$  is varied. For example, darker content can enjoy a higher acceptability than brighter content under the same scaling magnitude. Based on the sigmoid curve, we propose to employ a logistic function, an effective option for QoE modeling according to ITU-T [16], to model QoE versus scaling magnitude and luminance feature.

### 5.2 Logistic Regression Analysis

Logistic regression is a statistical analysis for binary response data whose probability of being positive is a sigmoid curve. This perfectly matches our data shown in Figure 5. The logistic function can be written as:

$$F(\vec{x}) = \frac{1}{1 + \exp^{-(\alpha + \beta x_1 + \gamma x_2 + \delta x_3 + \dots)}} \quad (3)$$

where  $\alpha, \beta, \gamma, \dots$  are the coefficients and  $\vec{x} = (x_1, x_2, \dots)$  are the predictor variables. In our user study, the predictor variables are the three stimuli, i.e., shot EMD, scaling magnitude of a shot, and device type. We use maximum likelihood estimation (MLE) to determine the coefficients. Data points that have studentized residuals less than -2 or greater than 2 are identified as outliers, which follows the standard logistic modeling in [30].

The modeling results demonstrate that all the predictors are significant except device type. Specifically, Wald test of device term,  $z = -1.313, p = 0.189$ , implies that the contribution of device type to model fit is not statistically significant. We also run likelihood ratio test to compare the two models with and without device type. The results,  $\chi^2 = 1.734$   $df = 1$   $p = 0.187$ , confirm that including device type into the model would not significantly improve model fit in a statistical sense. This is interesting as it is opposed to conventional video QoE for coding, where larger screen size relatively enlarges the coding artifacts and indeed deteriorates QoE. In the proposed scenario, however, people would always have an absolute perception of brightness regardless of screen size or pixel density. Considering that simplicity is always preferred under the same accuracy, we drop the device predictor and re-model the data using the other two predictors. We show the regression results in Table 2. We notice that all the coefficients are now statistically significant by passing Wald test and likelihood ratio test.

After model selection, we evaluate goodness of model fit. We carry out likelihood ratio test of model fit to compare our model with the perfect model that uses all degrees of freedom. The results,  $\chi^2 = 454.1$   $df = 1963$   $p > 0.999$ , is insignificant and indicate that there is very little unexplained variance and thus good model fit. We also compute multiple metrics of goodness of model fit in Table 3, i.e., root-mean-square error, Nagelkerke pseudo R-squared,

**Table 3: Model Evaluations**

	RMSE	$R^2$	$r$	$\rho$
Proposed	0.079	0.909	0.985	0.915
PSNR	NA	NA	0.798	0.803
SSIM	NA	NA	0.828	0.836

Pearson correlation coefficient ( $r$ ) and Spearman rank correlation coefficient ( $\rho$ ). The proposed model obtains a small RMSE and close-to-one values for the other three metrics, which implies an accurate and reasonable model.

We also compute the correlation between user experience data and objective metrics in conventional DBS, i.e., PSNR [11] and SSIM [8] between original and backlight-scaled images. The perceptual luminance of backlight-scaled images is calculated by the product of scaling magnitude and original pixel values. The PSNR and SSIM of a shot is the average over the frames within the shot. The results show that the proposed model is 9%~23% more correlated to the subjective data, which provides much more space to exploit user experience and enhance power performance.

Finally, we summarize the user acceptability model of scaling magnitude as

$$\mathcal{A} = \frac{1}{1 + \exp^{-(\alpha + \beta EMD_{shot} + \gamma \mathcal{M})}} \quad (4)$$

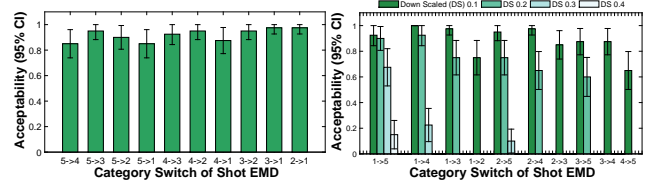
where the coefficients are shown in Table 2. Given an acceptability requirement, we would then be able to compute the best scaling magnitude of a given shot.

## 6. SMOOTHING DBS INCONSISTENCY

### 6.1 Motivational Study

We now study whether or not the continuous playback of backlight-scaled shots using the proposed scaling magnitude would cause flicker effect and how to smooth such, if any, temporal inconsistency. We proceed by studying the user acceptability of two continuously played shots because this can aid us to understand user experience at the minimum scaling unit. Thereby, we would be able to guarantee QoE for the whole video by individually smoothing the scaling at each shot boundary. Specifically, we aim to identify whether or not the magnitude of the second shot is proper given a known magnitude of the first shot. To this end, we set up a series of test clips of two consecutive shots (average duration 8 seconds), wherein the luminance feature  $EMD_{shot}$  of both first shot and second shot spans from category 1 to 5. We exclude the 5 clips whose first and second shots have the same luminance category because, under the same acceptability requirement, the optimal scaling magnitudes of these two shots will be very similar based on (4) and will show smooth variation.

We fix the scaling magnitude for the first shot  $\mathcal{M}_{fst}$  at the optimal value obtained from (4) using 0.9 acceptability. For the second shot of down-scaled clips ( $\mathcal{M}_{fst,opt} > \mathcal{M}_{sec,opt}$ ), we reduce the second magnitude  $\mathcal{M}_{sec}$  gradually from  $\mathcal{M}_{fst}$  at step of 0.1 until reaching the computed optimal magnitude of second shot, i.e.,  $\mathcal{M}_{sec} = \{\mathcal{M}_{fst} - 0.1, \mathcal{M}_{fst} - 0.2, \dots, \mathcal{M}_{sec,opt}\}$ . We do not include the versions whose second shot is scaled lower than  $\mathcal{M}_{sec,opt}$  since those second shots do not have the basic brightness to support user acceptability as shown in the scaling magnitude study. For example, suppose we have a clip with  $\mathcal{M}_{fst,opt} = 0.4$  and



**Figure 6: Acceptability of different patterns for up-scaled switch (left) and down-scaled switch (right).**

$\mathcal{M}_{sec,opt} = 0.7$  and we scale both shots at 0.4. Even though the backlight scaling is smooth, the second shot is too dark to be accepted because its scaling magnitude was already below 0.7 at the first place. Similarly, there is only one version for those up-scaled clips, i.e.,  $\mathcal{M}_{sec} = \mathcal{M}_{sec,opt}$ . Given these facts, we generate 30 versions of 20 test clips for this user study. Participants follow the same test protocol in scaling frequency test and a total of 1200 binary data points (400 up-scaled and 800 down-scaled) have been collected.

We discover an interesting distinction between the user experience of up-scaled and down-scaled clips. As shown in Figure 6, we see that the acceptability of all up-scaled clips are clearly high. This result is compatible with conventional QoE studies, where low-quality segments followed by high-quality segments was reported to be more preferred [31]. This phenomena might be due to the fact that users subconsciously believe brighter video is better and thus accept those up-scaled clips. Therefore, we decide to directly apply  $\mathcal{M}_{opt}$  for up-scaled shots. On the other hand, we observe that QoE of down-scaled clips decreases as the scaling variation increases. Furthermore, QoE under the same level of variation largely depends on luminance features. Considering these inter-related factors, we need to deliberately handle the down-scaled switches. In order to identify and smooth the inconsistent scaling, we propose to train a binary classifier to detect the inconsistency at every down-scaled switch and then smooth it by increasing  $\mathcal{M}_{sec}$  accordingly.

### 6.2 A Classifier for Smooth Scaling

Logistic regression is also a commonly used binary classifier [30]. Since it has been shown to accurately characterize the relation among QoE, content luminance and backlight, we build a logistic binary classifier to detect down-scaled inconsistency. We use the same method as shown in the previous section to process the outliers and fit the model.

Considering the stimuli of the motivational study, many possible features can be used to train the classifier: device type, variation of scaling magnitude  $\Delta\mathcal{M}$ , variation of luminance feature  $\Delta EMD_{shot}$ ,  $\mathcal{M}_{fst}$ ,  $\mathcal{M}_{sec}$ ,  $EMD_{shot,fst}$ , and  $EMD_{shot,sec}$ . We first exclude  $\mathcal{M}_{sec}$  and  $EMD_{shot,sec}$  because of their direct correlation to  $\Delta\mathcal{M}$  and  $\Delta EMD_{shot}$ , respectively. For example, given  $\mathcal{M}_{fst}$ ,  $\mathcal{M}_{sec}$  and  $\Delta\mathcal{M}$  can be mutually derived from each other. Besides, we find that the feature of device type does not pass Wald test ( $z = 1.672$   $p = 0.094$ ) and the likelihood ratio test of predictors ( $\chi^2 = 488.44$   $df = 765$   $p = 0.093$ ). This implies that device type is not a significant feature for the classifier. This is expected as we have discussed in the modeling of scaling magnitude. Finally, we obtain a classifier with 4 significant features as shown in Table 4.

To evaluate the classification performance of the classi-

**Table 4: Classifier Coefficients**

Feature	Estimate	Wald Test	Likelihood Ratio Test
Constant	-48.930	$z = -3.68$	NA
$\Delta EMD_{shot}$	0.088	$z = 7.05$	$\chi^2 = 848.99$ $df = 769$
$\Delta \mathcal{M}$	-54.775	$z = -10.00$	$\chi^2 = 517.80$ $df = 768$
$EMD_{shot,fst}$	-0.347	$z = -3.73$	$\chi^2 = 507.13$ $df = 767$
$\mathcal{M}_{fst}$	128.060	$z = 3.86$	$\chi^2 = 491.26$ $df = 766$

\*All tests are significant at  $p < 0.001$ .

**Table 5: Classifier Evaluations (%)**

	Precision	Recall	$F_1$
Proposed	87.79	93.09	90.36

fier, we perform 10-fold cross validation to data points. In Table 5, we show the precision, recall and F1 score ( $F_1 = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$ ). The results show that the classifier can achieve a reasonably good performance. Hence, we can rely on this classifier to detect the inconsistent down-scaled switches. Then the scaling magnitude of the detected second shot would be increased step by step (step size 0.05) until it can be accepted by the classifier.

## 7. SYSTEM IMPLEMENTATION

In this section, we demonstrate a practical application of the proposed DBS designs in a HTTP progressive streaming/download system to validate the benefits of QoE-driven power-saving strategy. We choose progressive streaming rather than adaptive streaming because we intend to keep the playing quality of videos unchanged, allowing a more accurate study of user experience in terms of DBS. We first summarize the proposed DBS profiler that addresses the issues in scaling frequency, magnitude and temporal consistency:

1. Determine the scaling boundary of the input video using shot detection via Algorithm 1.
2. Compute the best scaling magnitude for two consecutive shots under a given acceptability requirement using the logistic model in (4).
3. Detect the inconsistent down-scaled switch using the binary classifier and smooth the inconsistency, if needed, by gradually reducing the variation until it is accepted.
4. Iterate step 2 and 3 until all video shots are profiled.

We have implemented the full video-on-demand (VoD) system, where the university server stores encoded videos and users request the video using QoEPlayer. We consider two real-world scenarios. First, users access the video in the university using high-speed campus WiFi, which mimics the enterprise networks. Second, users view the video outside of campus using WiFi of third-party providers, which simulates the typical environment where videos are stored in the near CDN servers. In order to ensure that the QoE evaluation is only impacted by DBS, we choose not to select 3G/4G networks as their relatively unstable network condition and low data rate may cause networking-related distortions.

The server handles the computation of DBS and stores the DBS profile of a video as a separate file. Once a video is selected by the user, QoEPlayer will first download the corresponding DBS profile from the server if it is not already in the local storage and then proceed to the backlight-scaled

video playback. By offloading most DBS-related computations to the server side, the proposed DBS profiling can be efficiently treated by modern server clouds. Furthermore, considering that VoD services is time insensitive and a video only needs to be profiled once, we believe such an architecture is feasible and promising for power saving at the system level for massive number of devices.

The potential extra cost of the system lies in the transmission of the profiling file, which may introduce additional delay and power consumption. To investigate this issue, 5 participants are invited to watch clip “ducks” at both on-campus and off-campus environments. We ask them if they perceive any significant initial delay between two scenarios and how is the initial delay compared with their regular viewing. We are happy to report that none of them reports noticeable initial delay in both questions. In fact, this is expected since the file size of the profile is only around 1kB. Based on our measurement, it at most takes a few hundreds of milliseconds to download the file, which can hardly be perceived during the video start-up process. Hence, we conclude that the extra delay is negligible. We will show in the following that the additional power consumption is also marginal and does not compromise the benefits of the proposed designs.

We compare the performance of the proposed DBS designs with a conventional design, which uses the frame luminance variance [25] to decide scaling frequency and SSIM [8] to compute the scaling magnitude. The scaling magnitude of a video segment is computed by iteratively reducing the backlight (step size 0.05) until the minimum SSIM of the frames hits the given threshold. We combine these two schemes as they have been shown in prior sections to achieve better performance than existing approaches. The enhancement over such a strategy will highlight the advantages of the proposed designs. Both acceptability and SSIM requirement are set to be 0.9. The smoothing step size for reducing the down-scaled variation is set to be 0.05.

### 7.1 QoE Evaluations

We ask the participants to evaluate the acceptability of backlight-scaled videos using both the proposed profiling and the reference profiling. We focus on the on-campus scenario since we need to avoid the potential impact of video buffering on QoE. By using high-speed WiFi (around 4Mbps throughput), the video (bit-rate~2Mbps) can be play-backed uninterruptedly. Given that 4 full videos (average duration 134 seconds) are tested, we follow the test protocol as in the scaling frequency study.

As shown in Figure 7, the acceptability of the proposed DBS is higher than the reference designs for all clips. This is attributed to the fact that the proposed designs scale the backlight at appropriate time points with pleasing magnitude and smooth transition. In contrast, the reference designs cannot accurately capture the scaling time points and fail to smooth the temporal inconsistency of backlight changes. Although we notice that the average backlight level of videos using reference designs are higher than that of the proposed designs, this fact does not positively contribute to the overall QoE. This is because there is virtually no perceptual difference between the scaling magnitude enforced by the proposed designs and the reference designs. Thus, the high magnitude of reference designs is unnecessary.



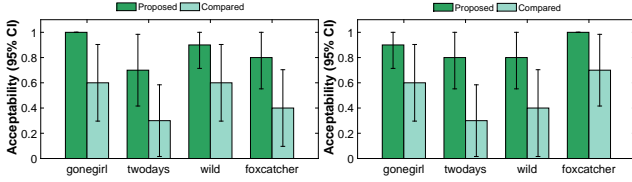


Figure 7: Acceptability on G Pro (left) and Nexus 4 (right).

## 7.2 Power Evaluations

In order to measure the battery power, we use Qualcomm Trepp, a measurement tool that directly reads hardware data from the devices. The test devices used in the system level evaluations, i.e., LG Optimus G Pro and Google Nexus 4, have been listed by Qualcomm as the supported devices that enjoy accurate power readings via Trepp [32]. During the system power measuring, we turn off all the unused apps and services. We also close all unnecessary networking activities, such as automatic update and Bluetooth, and only keep video transmission via WiFi active. We measure the power between the moment when users click a video and the moment video playback ends. After every measurement round, we make sure the cache is cleared and the entire video will be transmitted again in the next round.

We obtain the power reduction of different DBS designs based on the measurement method in [33]. First, we perform ten rounds of power measurement for a clip using three different strategies: (1) full backlight without DBS, (2) the proposed DBS, and (3) the reference DBS. Then the average power of a particular strategy can be computed. Finally, the power reduction of DBS over full-backlight playback is expressed as  $\frac{P_{full} - P_{DBS}}{P_{full}}$ , where  $P_{full}$  is the average power without DBS and  $P_{DBS}$  is the average power with DBS (either the proposed designs or the reference designs).

We show the results of power reduction in Figure 8. We observe that the proposed DBS designs accomplish a significant power savings. It can reach up to 42.34% power reduction depending on the device and content luminance. If a video is filmed in a darker environment and played on a larger screen, e.g., “wild” on G Pro, more power can be saved. The proposed DBS also outperforms conventional objective distortion based DBS by saving up to 20.95% more power when clip “wild” is tested on LG G Pro. The reason is that those objective distortion assumptions of conventional designs sacrifice much space for DBS and thus cannot maximally exploit the advantages of this technique. Although one may decrease the SSIM threshold to reduce the power of the reference designs, its QoE performance would be further downgraded, which is unacceptable for everyday applications. Therefore, we conclude that the proposed designs achieve best results in terms of both QoE and power.

## 8. DISCUSSION

**Luminance Compensation.** To absorb scaling distortion, several schemes have attempted to compensate the pixel luminance first and then scale the backlight. This luminance compensation is not adopted in this research because we would like to focus on the inherent correlation among QoE, backlight scaling, and content luminance at the funda-

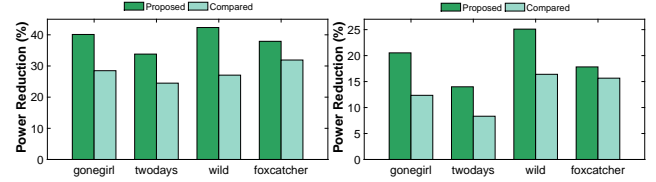


Figure 8: Power Reduction on G Pro (left) and Nexus 4 (right).

mental level. Adopting additional enhancement techniques, such as luminance compensation that relies on many objective or linear assumptions, into our study shall complicate the issues. Through luminance compensation, many image pixels can be fully compensated and how the resulting contrast distortion in the backlight-scaled image would impact QoE is unknown. It is true that luminance compensation is a complimentary technique to DBS. However, a separate full-scale study is definitely needed. Once the relationship between QoE and luminance compensation is better understood, we expect to achieve an even higher power saving.

**VoD System Design.** We have shown that the overhead of the extra profiling file is negligible. Alternatively, the profiling information can be easily embedded into the encoded video bitstream as supplemental enhancement information (SEI) of H.264 [34] for DBS at each mobile device to save energy. Such embedding as SEI can be seamlessly integrated into standard VoD systems. This will provide an elegant way to deliver such miniaturized yet essential information.

Note that both adaptive bit-rate streaming and backlight scaling have significant impacts on the overall QoE. They are two independent dimension of QoE. For example, in an adaptive streaming system, when one watches a video with frequently and abruptly scaled backlight, the overall QoE would be bad no matter how high the bit-rate is and how smooth the rate variation is. Although HTTP adaptive streaming is not currently supported in the proposed designs, this can be achieved easily by adopting a similar method to obtain the scaling magnitude and consistency models for different bit-rate/resolution. Similarly, DBS strategy for those rare video viewing scenarios, e.g., brighter outdoor or darker interior, can also be easily obtained.

## 9. CONCLUSION

In this paper, we have presented a novel approach for tackling the challenges of systematic power reduction on LCD smartphones from the perspective of QoE. We explore the QoE effects when dynamic backlight scaling is applied by conducting three motivational studies. Inspired by the interesting insights derived from these studies, we propose a suite of designs to address the DBS frequency, magnitude and temporal consistency. Through devising a VoD system that adopts the proposed designs, we demonstrate that the proposed DBS can achieve more than 40% system power savings, which represents a 20% improvement over conventional non-QoE schemes, without QoE impairment.

We would like to emphasize that this new approach is fundamentally different from conventional video QoE research which mainly aims to improve encoding and networking performance. This research is indeed a pioneering step towards exploring QoE to enhance power performance. Especially,

this approach extends a promising dimension for QoE studies. Furthermore, this research represents a significant attempt to address the prevailing issue of mobile display power in the application layer. The preliminary results have shown that such interdisciplinary research can achieve much desired display power performance in contemporary .

## 10. ACKNOWLEDGEMENTS

This research is supported by NSF Grants ECCS-1405594 and ECCS-1406154.

## 11. REFERENCES

- [1] A. Carroll and G. Heiser. An analysis of power consumption in a smartphone. In *USENIX annual technical conference (USENIXATC)*, June 2010.
- [2] DisplaySearch. <http://goo.gl/b56ebH>.
- [3] W.-C. Cheng and M. Pedram. Power minimization in a backlit tft-lcd display by concurrent brightness and contrast scaling. *IEEE Trans. Consum. Electron.*, 50:25–32, February 2004.
- [4] N. Chang, I. Choi, and H. Shim. Dls: dynamic backlight luminance scaling of liquid crystal display. *IEEE Trans. VLSI Syst.*, 12:837–846, August 2004.
- [5] A. Iranli, H. Fatemi, and M. Pedram. Hebs: histogram equalization for backlight scaling. In *Proc. of Design, Automation and Test in Europe (DATE)*, March 2005.
- [6] A. Iranli, W. Lee, and M. Pedram. Hvs-aware dynamic backlight scaling in tft-lcds. *IEEE Trans. VLSI Syst.*, 14:1103–1116, October 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 international conference on Embedded software (EMSOFT)*, October 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.*, 63:335–348, February 2014.
- [9] Y. Liu, M. Xiao, M. Zhang, X. Li, M. Dong, Z. Ma, Z. Li, and S. Chen. Content-adaptive display power saving in internet mobile streaming. In *ACM NOSSDAV*, March 2015.
- [10] P. Ni, R. Eg, A. Eichhorn, C. Griwodz, and P. Halvorsen. Flicker effects in adaptive video streaming to handheld devices. In *ACM MM*, November 2011.
- [11] 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*, 2007(1):4–4, 2007.
- [12] W. Lin and C.-C. Kuo. Perceptual visual quality metrics: a survey. *Journal of Visual Communication and Image Representation*, 22:297–312, May 2011.
- [13] W. Song and D. Tjondronegoro. Acceptability-based qoe models for mobile video. *IEEE Trans. Multimedia*, 16:738–750, April 2014.
- [14] F. Zhang, W. Lin, Z. Chen, and K. N. Ngan. Additive log-logistic model for networked video quality assessment. *IEEE Trans. Image Process.*, 16:1536–1547, April 2013.
- [15] C. Alberti, D. Renzi, C. Timmerer, C. Mueller, S. Lederer, S. Battista, and M. Mattavelli. Automated qoe evaluation of dynamic adaptive streaming over http. In *IEEE QoMEX*, July 2013.
- [16] Opinion model for video-telephony applications. ITU-T Recommendation G.1070. 2007.
- [17] W. Song, D. Tjondronegoro, and M. Docherty. Saving bitrate vs. pleasing users: where is the breakeven point in mobile video quality. In *ACM MM*, November 2011.
- [18] H. Knoche, J. McCarthy, and M. Sasse. Can small be beautiful?: assessing image resolution requirements for mobile tv. In *ACM MM*, November 2005.
- [19] GT Fechner. Element of psychophysics. *New York: Holt, Rinehart and Winston*, 1966.
- [20] A. Balachandran, V. Aggarwal, E. Halepovic, J. Pang, S. Seshan, S. Venkataraman, and H. Yan. Modeling web quality-of-experience on cellular networks. In *ACM MobiCom*, September 2014.
- [21] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang. Developing a predictive model of quality of experience for internet video. In *ACM SIGCOMM*, August 2013.
- [22] Xiph. <https://media.xiph.org/video/derf/>.
- [23] DisplayMate. Smartphone displays under bright ambient lighting shoot-out. <http://goo.gl/n89Kgo>.
- [24] W. Song, D. Tjondronegoro, and I. Himawan. Acceptability-based qoe management for user-centric mobile video delivery: a field study evaluation. In *ACM MM*, November 2014.
- [25] S. Pasricha, M. Luthra, Shivajit Mohapatra, N. Dutt, and N. Venkatasubramanian. Dynamic backlight adaptation for low-power handheld devices. *IEEE Des. Test. Comput.*, 21:398–405, September 2004.
- [26] R. Lienhart. Comparison of automatic shot boundary detection algorithms. In *SPIE Electronic Imaging*, pages 290–301, 1998.
- [27] A. Hanjalic. Shot-boundary detection: unraveled and resolved? *IEEE Trans. Circuits Syst. Video Technol.*, 12:90–105, February 2002.
- [28] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover’s distance as a metric for image retrieval. *International journal of computer vision*, 40:99–121, November 2000.
- [29] A M. Ferman, A M. Tekalp, and R. Mehrotra. Robust color histogram descriptors for video segment retrieval and identification. *IEEE Trans. Image Process.*, 11:497–508, May 2002.
- [30] R. Christensen. *Log-Linear Models and Logistic Regression*. Springer-Verlag, New York, USA, 1997.
- [31] M. Zink, O. Künzel, J. Schmitt, and R. Steinmetz. Subjective impression of variations in layer encoded videos. In *IEEE/ACM IWQoS*, June 2003.
- [32] Qualcomm. Devices that report accurate battery power. <http://goo.gl/ZecX96>.
- [33] M. Dong and L. Zhong. Power modeling and optimization for oled displays. *IEEE Trans. Mobile Comput.*, 11:1587–1599, September 2012.
- [34] T. Wiegand, G.J. Sullivan, G. Bjontegaard, and A. Luthra. Overview of the h.264/avc video coding standard. *IEEE Trans. Circuits Syst. Video Technol.*, 13:560–576, July 2003.