IBC2023

CONTENT-DEPENDENT POWER SAVING MODEL FOR HDR DISPLAY DEVICES

Jayasingam Adhuran, Gosala Kulupana

British Broadcasting Corporation, United Kingdom

ABSTRACT

Widespread adoption of High Dynamic Range (HDR) videos elevates the in-home experience of video consumption. However, displaying HDR video content can escalate the power consumption of the TVs to over 300W, a figure that is content-dependent. Moreover, existing solutions adversely impact visual fidelity in the attempt to reduce power consumption. In response, this paper proposes a just noticeable difference model comprising regions of interest detection, luminous adaptation and spatial correction techniques to conserve power. The model also incorporates skin detection and visual information fidelity based optimization techniques to reduce the visual fidelity loss. Extensive experiments conducted on multiple modes of LCD and OLED TVs demonstrate significant savings achieving an average of 1-18% power reduction. The best performing variant of the proposed JND model can achieve an average power reduction of 41W and up to 69W with LCD cinema home mode.

INTRODUCTION

End-user Quality of Experience (QoE) has increased dramatically in recent years with the introduction of High Dynamic Range (HDR) with high-resolution video formats such as Ultra High Definition (UHD) [1] [2]. As such, video streaming providers have shown great interest in delivering HDR services to customers. Consequently, the world has seen a rapid proliferation of advanced display devices such as televisions (TVs) and mobile phones supporting the new video technologies that elevate the in-home experience of video consumption. In this trend, Light Emitting Diode (LED) display technologies have flourished in the past decade, superseding traditional Liquid Crystal Displays (LCD)s with fluorescent backlights, owing to improvements in multiple aspects such as brightness, visual fidelity and power savings. Organic LED (OLED) displays, a variant of the LED technologies, are widely used within high-end consumer devices to provide enhanced image quality [3].

HDR imaging delivers an increased range of luminance, colour gamut and contrast, which shows significant improvements over Standard Dynamic Range (SDR) formats [4] [5]. However, displaying HDR videos comes at an increased cost of power consumption despite the mitigation measures that OLED displays provide. Although standard average power consumption is expected to be around 120W, it may not apply to HDR videos. Power usage for displaying HDR videos is content-dependent and shows great variation from one video to another. Consequently, some, HDR video contents can escalate the power consumption to over 300W for certain video sequences.

Literature in video compression often narrates the deployment of Just Noticeable Difference (JND) models [6] [7]. More often they are used in applications that require the exploitation of perceptual redundancy. In general, JND models inject contaminations to images and videos up to a limit where perceptual differences between the original and the contaminated

are minimized. The decomposition of the images plays a major role in a JND model in order to identify perceptual redundancy. In this context, various algorithms have been proposed in the past [8] [6] [9]. Importantly, the utilization of JND models has also been explored in the power reduction of displaying images and videos in OLED displays [10] [11].

The major drawback of existing solutions is the detrimental effect on the visual fidelity. Moreover, literature has not explored power reduction in HDR video content. To this end, a JND model is presented in this paper that is capable of reducing power requirements to display HDR video at a minimal loss of visual fidelity. The proposed JND model leverages deep learning based Regions Of Interests (ROI) detection, luminous adaptation and spatial correction to generate a mask to contaminate the source video. The major contributions of this research are 1. ROI based JND model that can reduce the power requirements of HDR videos at minimal fidelity loss; 2. power and visual quality related analysis in the context of HDR videos, OLED and LCD displays. It is anticipated that the proposed technology would operate at the decoder-side (i.e., TVs, set-top boxes, mobile phones) with a separate mode that would allow the technology to be enabled or disabled as per the user's discretion.

The rest of the paper is organized as follows. Firstly, existing works are discussed. Next, the overview of the proposed model and individual components are elaborated in the methodology section. Then, the experimental procedure and the results are reported and discussed before presenting the concluding remarks in the final section.

RELATED WORK

The power consumption of OLED displays can be modelled as a proportion to the sum of power consumed by each pixel of a video frame [12] [13] [14]. The power consumption also accounts for the supplementary power demand required to drive the display including the display safety measures. Exploiting pixel-centric characteristics of OLED displays, multiple research works exploit the image characteristics to reduce either intensity or the luminance level of the images [1] [15] [3] [16] [17] [10].

Chondro et al. [1] proposed a dimming technique incorporating power law based pixel transformation, scene detection and hue preservation technique. Although up to 73% power saving was reported, the fidelity is greatly affected resulting in not more than 0.586 in Visual Information Fidelity (VIF) [18]. Shin et al. [17] proposed a deep learning and power constraints incorporated framework built on top of a conditional generative adversarial network to influence the contrast of an image. Although results indicate a high degree of Visual Saliency-induced Index (VSI) [19] score, the fidelity of the test images and videos remains heavily affected. Saliency-modulated JND model has been used to save energy from images [10]. Here saliency maps developed in the block-based Discrete Cosine Transform (DCT) domain are applied to the images before subjecting them to perceptual quality optimization. Applying statistical analysis on a binary voting system by subjects, the research work concluded that 80% of the test images did not have a major difference in perceptual quality. Although 14% energy saving is also reported for the test conducted on a 12" OLED display, the major contribution has been produced from the images that failed the statistical tests. Moreover, an ROI based pixel and subpixel manipulation presented in [20] report around 19% power savings at 10-12% brightness reduction and colour degradation. Moreover, experiments in the existing research works have been limited to SDR videos.

To primarily address the concerns related to visual fidelity loss, the proposed JND model, exploits useful components from existing literature, adopts the fundamental norm of pixel intensity reduction and integrates deep learning and image processing tools resulting in power savings at a minimal loss to visual quality. Moreover, the experiments are conducted

in the context of HDR sequences to resolve the shortcomings of HDR sequences in this scope of research.

METHODOLOGY

Overview of the proposed framework

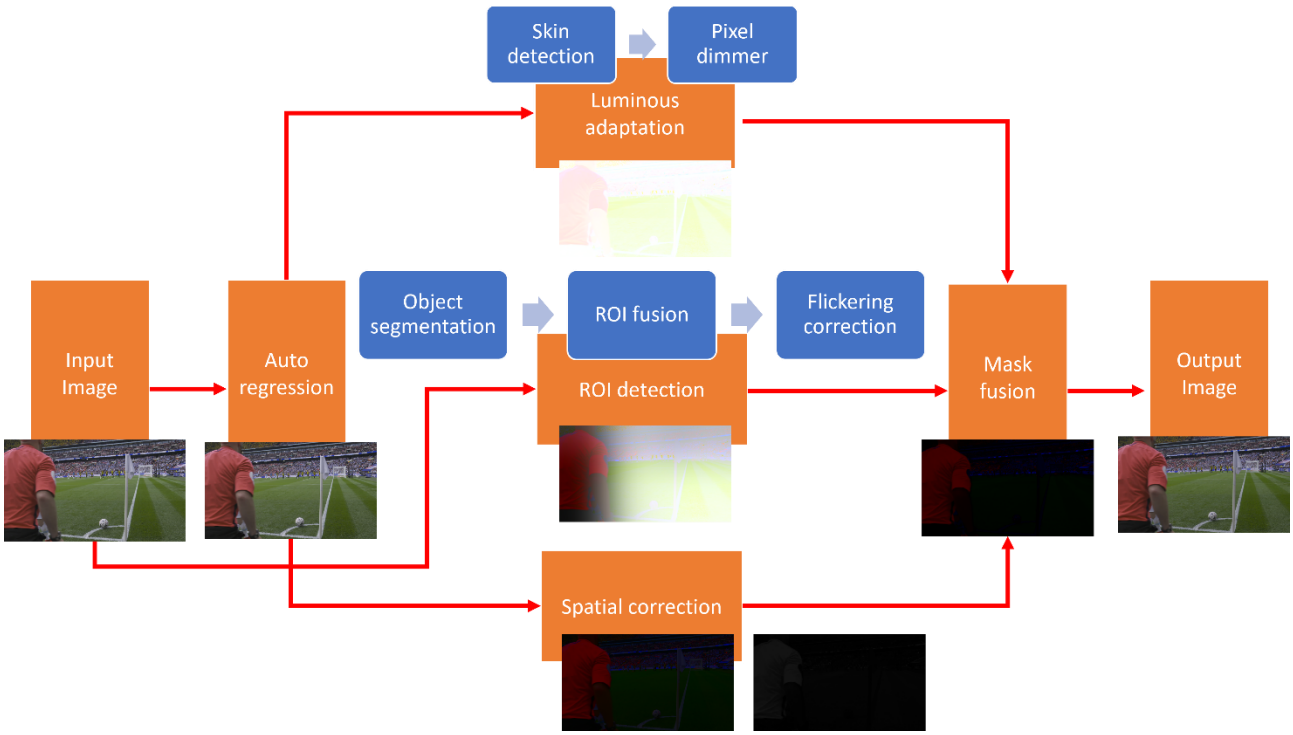


Figure 1 - Framework of the proposed JND model

Figure 1 demonstrates the framework of the proposed JND model. A luma signal of the input image $I_{(x,y)} \in \mathbb{R}^{W \times H}$ is first sent to the autoregression module. The autoregression module used here is inspired by the study [21] that proposed JND estimation using free energy principle assuming that Human Visual System (HVS) perceives orderly contents. The autoregression module applies a decomposition technique that eventually predicts the pixels based on the neighbouring pixels. The luminous adaptation module applies background intensity reduction on the image $I^{rec}_{(x,y)} \in \mathbb{R}^{W \times H}$ that is retrieved from the autoregression module. It also comprises hue characterized skin detection technique in order to preserve the hue degradation in identified areas. The spatial correction module applies a series of filters such as Gaussian and box as well as gradient filters on $I^{rec}_{(x,y)} \in \mathbb{R}^{W \times H}$. Additionally, the variation between such filtered images and $I_{(x,y)} \in \mathbb{R}^{W \times H}$ is also obtained. This results in the removal of several pieces of information, preserving sharp transitions and mean variations from the images. The ROI detection module first implements a pre-trained maskRCNN model for object segmentation based on image $I_{(x,y)} \in \mathbb{R}^{W \times H}$. The segmented objects are fused in a non-linear manner in order to preserve spatial correlation that aids encoding processes. Since the spatial location of the object mask can vary in the subsequent framing in the presence of motion sequences, temporal Gaussian masking is applied to smoothen the variation across frames to avoid any flickering artefacts. The mask generation module fuses the respective outputs from luminous adaptation, spatial correction and ROI generation modules to generate the JND mask $JND_{o,t}$. In this context, a VIF based

optimization process is followed. The generated JND mask is also influenced by the mask of the previous frame to maintain consistency between successive frames. The JND masked is normalized and applied to the image $I_{(x,y)} \in \mathbb{R}^{W\times H}$ to generate the JND contaminated output image $I^{o}_{(x,y)} \in \mathbb{R}^{W\times H}$

Algorithm 1: Deduction of predicted image.

Input: Image $I_{(x,y)}$, constant matrix $\mathbb{K}^{7\times7}$, $var_{(x,y)}$, R, r

Output: $I^{dec}_{(x,y)}, w, w_{max}$

$I^{pad}_{(x,y)} \leftarrow$ Pad image $I_{(x,y)}$ within $[R+r, R+r]$

$I^{o}_{(x,y)} \leftarrow$ Truncate $I^{pad}_{(x,y)}$ by R pixels in all direction

Initialize $I^{dec}_{(x,y)}, w, w_{max}$

for $u, v \in [-R, R+1], u, v \neq 0$

$I^{move}_{(x,y)} \leftarrow$ Shift $I^{o}_{(x,y)}$ by u, v pixels on $I^{pad}_{(x,y)}$ in both directions.

$I^{dec}_{(x,y)} \leftarrow$ Truncate $I^{move}_{(x,y)}$ by r pixels in all directions.

$d = \frac{-\left(\left(I^{o}_{(x,y)} - I^{move}_{(x,y)}\right)^2 * \mathbb{K}\right)}{var_{(x,y)}}$

$I^{dec}_{(x,y)} \leftarrow I^{dec}_{(x,y)} + I^{dec}_{(x,y)} \times e^{d}$

$w \leftarrow w + d$

$w_{max} \leftarrow d$, if $d > 0$

$w \leftarrow w + w_{max}$

$w \leftarrow 1$, if $w = 0$

Autoregression

The autoregression module retrieves a predicted image $I^{rec}_{(x,y)}$ from a source image $I_{(x,y)}$ by predicting pixels from the neighbouring pixels in order to remove non-orderly components from the image [21]. Given an image $I_{(x,y)}$,weight w and maximum weight w_{max} that are needed to generate $I^{rec}_{(x,y)}$ can be deduced as shown in Algorithm 1 and where $\mathbb{K}^{7\times7}$, $var_{(x,y)}$, R and r are kernal of size 7×7, image variance that accounts for both pixel intensities and background luminance. $var_{(x,y)}$ can be given by,

$$var_{(x,y)} = \begin{cases} {\sigma_b}^2, & \sigma_I < \sigma_b \\ \left(\sigma_b \times \sqrt{\frac{\sigma_b}{\sigma_I}}\right)^2, & otherwise \end{cases}$$

where, $\sigma_b = b_g\left(S\left(I_{(x,y)}\right)/32\right)$ and $\sigma_I = \sqrt{I^2_{(x,y)} * \mathbb{K}^{7\times7} - \left(I_{(x,y} * \mathbb{K}^{7\times7}\right)^2}$. Here, $S(\cdot)$ refers to the spatial gradient filtering process that extract background luminance. Moreover, b_g is a luminous modulation operator [21] adopted to support 10-bit picture which is given using k as,

$$b_g(k) = \begin{cases} 68 \times \left(1 - \sqrt{\left(\frac{k-1}{512}\right)}\right) + 12, & k < 512 \\ \frac{3}{128} \times (k - 512) + 12, & otherwise \end{cases}$$

Then, the predicted image $I^{rec}_{(x,y)}$ is deduced from the following relationship.

$$I^{rec}_{(x,y)} = \begin{cases} \frac{I_{(x,y)}}{w}, & I^{dec}_{(x,y)} = 0 \\ \frac{I^{dec}_{(x,y)} + I_{(x,y)} \times w_{max}}{w}, & otherwise \end{cases}$$

Luminous adaptation

The luminous adaptation module consists of an independently processed skin detection technique and a background pixel dimmer technique.

Skin detection

Skin detection requires the selection of appropriate colour space [22]. In that context, $I_{(x,y)}^{rec}$ with its chroma components from the source image is converted to the 8-bit HSV colour space. By applying lower and upper ranges of hue, saturation and value to (0,180), (0,38), (120, 255) a mask JND_s can be generated that can filter out skin detected regions at the pixel level. Here, the aforementioned threshold ranges are selected empirically. Furthermore, JND_s is normalized between 1 and ϑ, a user defined luminous adaptation parameter.

Pixel dimmer

A mask JND_{pd} required to suppress the pixel intensity from the background regions is generated from $I_{(x,y)}^{rec}$ by subjecting it to the Gaussian $G(\cdot)$ processes and luminous modulation operator b_g as given by the following equation.

$$JND_{pd} = b_g\big(G\big(I_{(x,y)}\big)/32\big)$$

Spatial correction

The spatial correction module aims to preserve edges, lines and mean variation of the pixels and generate two independent masks JND_{el} and JND_{mv}. In conjunction with the ROI mask, the edges could be filtered or blurred out. To tackle this problem and to reduce the mean pixel variations of the pixels, the masks are deployed independently. JND_{el} and JND_{mv} can be given by,

$$JND_{el} = abs\big((a \times S\big(I_{(x,y)}\big)/32 + b) \times l_g + 0.5 - c \times S\big(I_{(x,y)}\big)/32\big)$$

$$JND_{mv} = abs\left(\frac{I_{(x,y)} - 0.5 \times \frac{\sum I_g}{n}}{0.5 \times \frac{\sum I_g}{n}}\right)$$

where $l_g = t \times \log\left(1 + \frac{C\big(I_{(x,y)}\big)}{t}\right)$ and $C(\cdot)$ represent series on 2-D convolutions with one gradient, two box and one Gaussian kernel of size 5×5. a, b and c are empirically driven parameters.

ROI detection

In ROI detection module, a pre-trained maskRCNN [23] model is deployed for object segmentation. The maskRCNN is a deep learning framework that has been developed on top Region Proposal Network (RPN) which employs a neural network to identify ROIs based on binary classifiers. The mask RCNN module used in this research deploys Resnet 50 and Feature Pyramid Network (FPN) as the backbone (Resnet-50-FPN) and has been pre-trained on COCO dataset with 80 classes [24]. Image $I_{(x,y)}$ is converted RGB colour space along with its chroma components and subsequently resized to 1280×720 resolution before applying the model. The aforementioned resolution chosen such that a tradeoff between complexity and picture quality can be achieved. Thereafter, the detected segments are concatenated using bitwise operations to generate an ROI mask JND_{roi}. However, an ROI mask in its nominative form could adversely affect the encoding process disrupting the spatial correlations between pixels. In order to mitigate that, a logarithmic filter is used to

alleviate the sudden transition of pixel intensity at the ROI boundary. Then, JND_{roi} can be defined as ,

$$JND_{roi} = G\left(I_{(x,y} \times \left(\frac{1}{1-\alpha \times \log\left(\frac{I_{(x,y}}{max(I_{(x,y)})}\right)}\right)\right)$$

where α is a design parameter derived empirically. Moreover, JND_{roi} is normalized between 0 and 1 so that it can be used with JND_{mv} when fusing the masks. Applying ROI mask can introduce flickering artefacts. Therefore, the temporal Gaussian filter with $\sigma = 10$, incorporates three past frames to smoothen the sudden intensity variation along the temporal domain. Subsequently, δ is introduced to control the effect of JND_{roi} on final mask $JND_{o,t}$.

$$JND_{roi} = clip(\delta \times JND_{roi}, 0, 1)$$

Thereafter, the masks are cached for future use.

Mask fusion

All masks generated are fused to develop the final JND mask. However, determining the fusion parameters is challenging. An optimization process is used to identify the fusion parameters subject to achieving a user-specified VIF threshold v_t. By applying gradient-based minimization process, the following objective function $\mathcal{J}(\rho)$ can be optimized.

$$\mathcal{J}(\rho) = clip\left(\mu \times JND_{el} + \rho \times JND_{roi} \times JND_{mv} + \vartheta \times JND_{pd}, 0, 1023\right)$$

subject to: $$clip(v_t + 0.1, v_t, 1.0) \geq vif\left(I_{(x,y}\right) \geq v_t$$

Here μ and ϑ are empirically tuned parameters; ρ is obtained from the optimization process and $vif(\cdot)$ denotes VIF calculations. Moreover, the optimization function does not involve skin detection mask JND_s in order to maintain fidelity without a specific focus on skin tone. Moreover, not more than 10 steps of iterations would be allowed for the optimization process in order to reduce the complexity. Post determination of ρ, the final JND mask $JND_{o,t}$ for a given frame t can be derived as follows.

$$JND_{o,t} = clip\left(\mu \times JND_{el} + \rho \times JND_{roi} \times JND_{mv} + JND_s \times JND_{pd}, 0, 1023\right)$$

A masking correction is applied based on the JND mask of the previous frame $JND_{o,t-1}$as a second stage of flickering artefact mitigation. Then, the revised JND mask can then be derived as,

$$JND_{o,t} = clip\left(\left(JND_{o,t} - \varepsilon \times G\left(JND_{o,t} - JND_{o,t-1}\right)\right), \epsilon^-, \epsilon^+\right)$$

where $\epsilon^- = minimum\left(floor\left(JND_{o,t-1}\right), 0\right)$ and $\epsilon^+ = maximum\left(ceil\left(JND_{o,t-1}\right), 1023\right)$. Moreover, the aforementioned correction is not applicable to the first frame. Subsequently $JND_{o,t}$ is inversely normalized between 1 and 0. Finally the output image $I^o_{(x,y)}$ can be obtained.

$$I^o_{(x,y)} = JND_{o,t} \times I_{(x,y)}$$

EXPERIMENTS AND RESULTS

The proposed JND masks were applied to the test sequences with the following parametric values $\alpha = 0.2, \delta = 0.5, \varepsilon = 0.2$ and $\mu = 0.5$. Moreover, the user-specified VIF threshold was assumed to be $v_t = 0.8$, a value that indicates high levels of visual fidelity and much greater

than Chadro et al. could achieve [1]. Furthermore, variants of the proposed JND model were constructed by varying ϑ and including/excluding skin detection from the proposed JND model. In this context, four variants experimented with are listed below.

- JND Var 1: $\vartheta = 2.5$ and $JND_{sd} = 1, \forall(x, y)$
- JND Var 2: $\vartheta = 2.5$
- JND Var 3: $\vartheta = 1.5$ and $JND_{sd} = 1, \forall(x, y)$
- JND Var 4: $\vartheta = 1.5$

Further, static pre-scaling parametric models were also implemented for comparison with the proposed JND models. In this context, gamma, contrast and saturation are slightly modified to contaminate the source video. FFMPEG [25] library is used to introduce the parametric adaptation to source videos and the following variants are obtained empirically such that power can be reduced at a considerable level of visual fidelity on OLED displays.

- Pre-scale 1: gamma = 0.7, contrast = 0.7, saturation = 0.9
- Pre-scale 2: gamma = 0.9, contrast = 0.9, saturation = 0.9

The parameter selection of Pre-scale 1 ensures heavy power reduction whereas that for Pre-scale 2 matches the fidelity of the proposed JND models.

To evaluate the performance, tests are performed on fifteen BBC HDR Hybrid Log-Gamma (HLG) clips extracted from FA cup (sports), His Dark Materials (HDM) (drama) and Nat His (Natural History). Each category comprises 5 videos that are 30 seconds long and 3840x2160 in resolution. The source and the contaminated videos were encoded using x265 encoder with a target bitrate of 18 Mbps. Thereafter, the videos were converted to the Apple ProRes 422 format in order to be able to play the videos into the TVs using an external source. The purpose of using an external device was to nullify the power requirements to decode the videos. Moreover, three 65 inch commercial TVs were chosen as the display devices: Sony Bravia LCD (LED backlight), LG OLED C2 and LG OLED B1. Also, two modes namely standard and filmmaker from OLED TVs and standard and cinema home from LCD TV were selected for experiments. The TVs were directly powered by Bryant power distribution unit which also aids the acquisition of power readings.

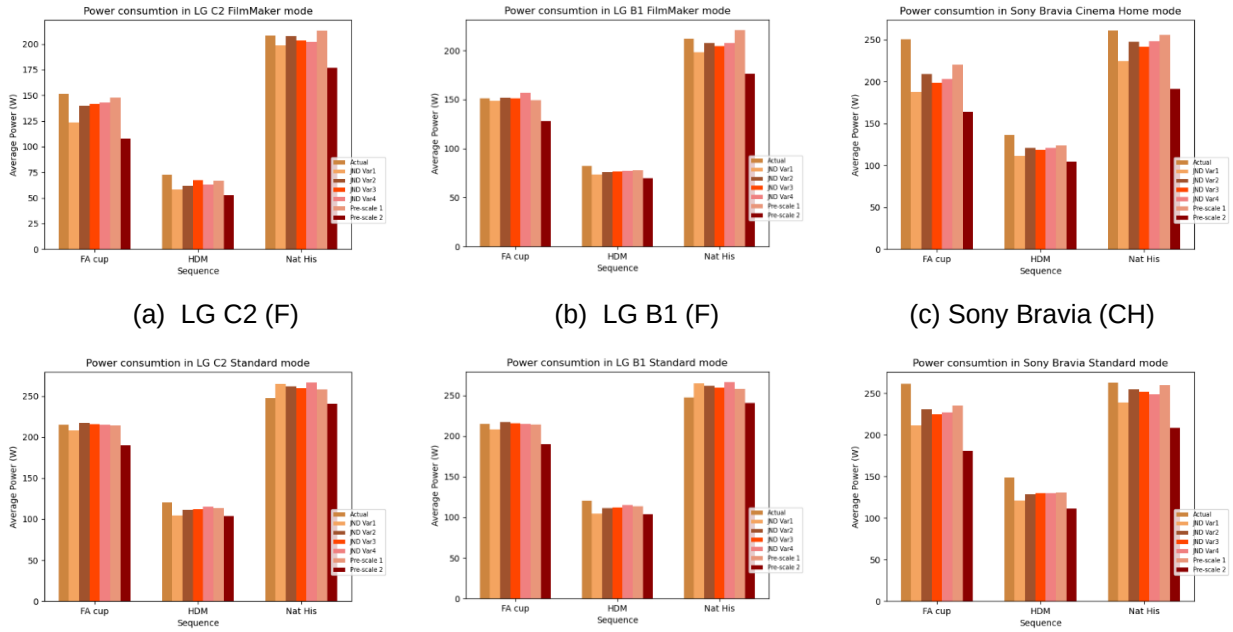
Evaluation metrics

Performance needs to be assessed as a trade-off between power savings and visual fidelity. The amount of power saved can be measured by the difference $(\Delta P = P_s - P_c)$ between the power consumed for displaying the source and the contaminated videos. Here, P_s refers to the power consumed by the source video whereas P_c denotes power used by the contaminated videos. Moreover, a relative measure of power saved can be measured as a percentage as $\%P = \frac{P_s - P_c}{P_s}$. Additionally, the average pixel intensity difference as a percentage of the source pixel intensity across the video is also presented in this paper. Furthermore, there is no standard quantitative measure for the objective quality assessment of the contaminated videos. Therefore, in this research, VSI and VIF have been adopted to evaluate the objective quality as they have been used in similar research works [1], [17].

Performance evaluation

The performance of the proposed variants of the JND models is presented in Figure 2 (zoom in for better clarity) and Table 1. For simplicity, the average power consumption from the five videos in each category is presented in Figure 2. Here, F, S and CH denote filmmaker, standard and cinema home modes respectively. As expected, there is a considerable amount of power saving for the utilization of LCD screen in both cinema home and standard

modes. Conversely, in OLED displays, moderate power savings can be obtained in filmmaker modes. Moreover, power reduction is unlikely in OLED standard modes with the only exception being the HDM sequences. Individual analyses of each video indicated that power saving and loss can vary from one sequence to another and they are also deeply dependent on the display used. Furthermore, it is evident from Table 2 that the best performing variant, JND var 1 makes the maximum power savings among the proposed JND models and yields up to 69W (for a FA cup clip with Sony Bravia CH mode) as a result of higher luminous adaptation parameter ϑ. Moreover, the reduction of ϑ and the incorporation of skin detection reduces the power savings. In this line of trend, JND var 2, JND var 3, and JND var 4 produce loss in LG B1 standard mode owing to the heavy power consumption by the Nat His sequences which are bright in nature. While ϑ remains the dominant factor for power savings in LCD displays, face detection masking also provides a significant contribution to the power saving as opposed to ϑ in LG C2 filmmaker mode at a lower ϑ.



(a) LG C2 (F) (b) LG B1 (F) (c) Sony Bravia (CH)

(d) LG C2 (S) (e) LG B1 (S) (f) Sony Bravia (S)

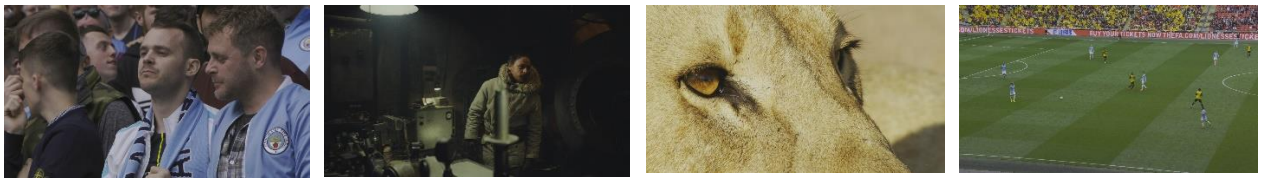
Figure 2 - Average power consumption of grouped video sequences on respective displays and modes.

Table 1 - Performance analysis of the variants of the proposed JND model

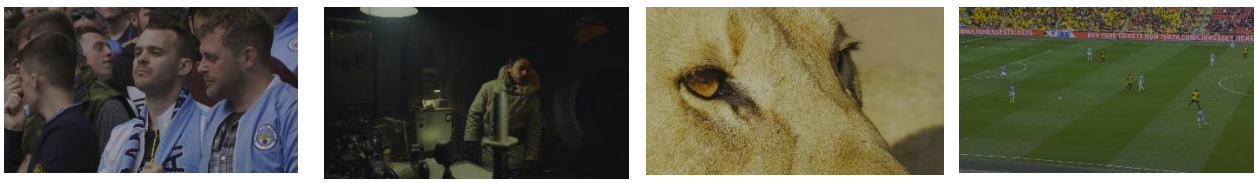
Metric	Display type	JND Var 1	JND Var 2	JND Var 3	JND Var 4	Pre-scale 1	Pre-scale 2
ΔP (W)	LG C2 (F)	17.45	7.69	6.78	8.33	1.62	31.71
$\%P$		14.14	7.35	4.74	7.20	2.72	23.28
ΔP (W)	LG C2 (S)	11.66	4.99	5.89	1.58	2.49	25.11
$\%P$		9.79	5.12	5.38	2.21	3.31	17.32
ΔP (W)	LG B1 (F)	8.20	3.08	4.08	1.01	-0.91	23.73
$\%P$		5.91	2.61	2.96	1.00	0.58	15.42
ΔP (W)	LG B1 (S)	1.68	**-2.56**	**-1.59**	**-4.52**	-0.93	16.05
$\%P$		2.93	0.08	0.34	**-1.28**	0.42	9.20
ΔP (W)	Sony Bravia (CH)	41.28	23.59	29.77	25.13	16.14	62.83
$\%P$		18.88	10.98	13.55	11.55	7.76	27.96

ΔP (W)	Sony Bravia (S)	33.86	19.71	21.98	22.17	15.95	57.41
$\%P$		15.45	9.40	10.06	10.11	7.83	25.29
VSI		0.99	0.99	0.99	0.99	0.99	0.99
VIF		0.87	0.86	0.87	0.82	0.84	**0.66**
Minimum VIF		0.75	0.74	0.77	0.73	0.74	**0.55**
Intensity difference (%)		11.51	12.26	12.16	18.61	5.39	21.25

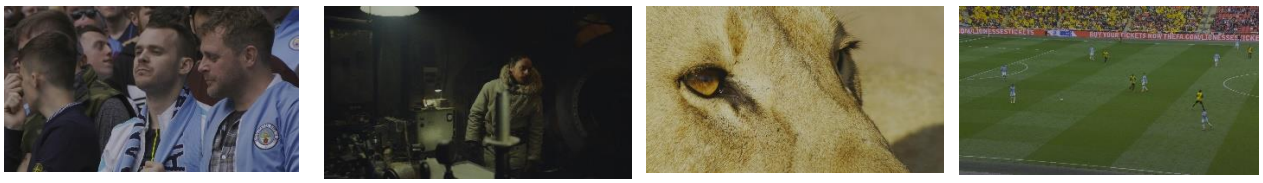
Quantitative analysis of the visual quality levels indicates that VSI remains constant for all variants tested including the pre-scale variants. However, VIF scores remain high for the proposed JND models and Pre-scale 1 with minimum VIF is well above 7.0. Minimum VIF is critical as few frames with very low fidelity could adversely affect the end-user experience. Although Pre-scale 2 demonstrate very high power savings, it comes at a very high cost of the quality. For further visual inspection, the first frame of selected video clips is illustrated in Figure 3 (they have not been corrected for print media, so are best viewed on an HDR display). Although there exists a perceptual difference between the source and the proposed variants, the fidelity of the proposed variants remains high in contrast to the Pre-scale 1 which deteriorates the visual fidelity.



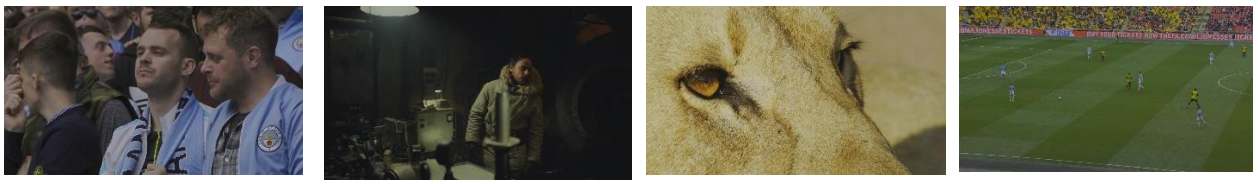
(a) source



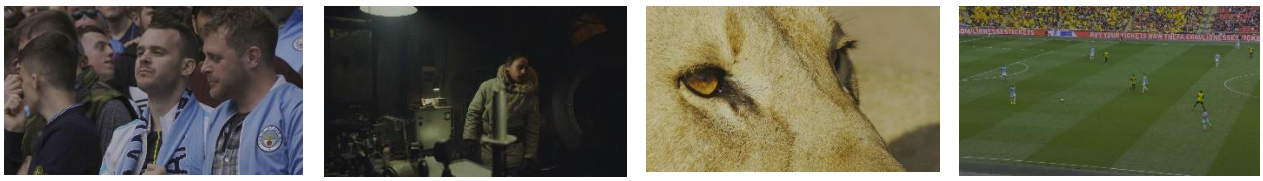
(b) JND var 1 (ΔP for LG C2 (F) = 23.05W, 13.31W, 32.00W, 41.83W)



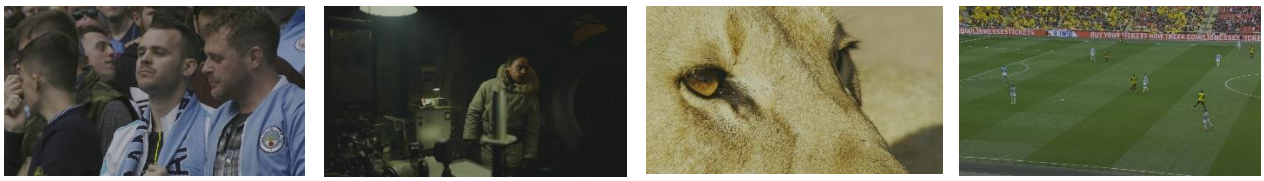
(c) JND var 2 (ΔP for LG C2 (F) = 13.63W, 10.47W, 19.08W, 16.49W)



(d) JND var 3 (ΔP for LG C2 (F) = 9.90W, 1.11W, 19.52W, 17.32W)



(e) JND var 4 (ΔP for LG C2 (F) = 7.80W, 6.25W, 23.51W, 6.69W)



(f) Pre-scale 1 (ΔP for LG C2 (F) = 12.54W, 4.80W, 9.27W, 7.71W)

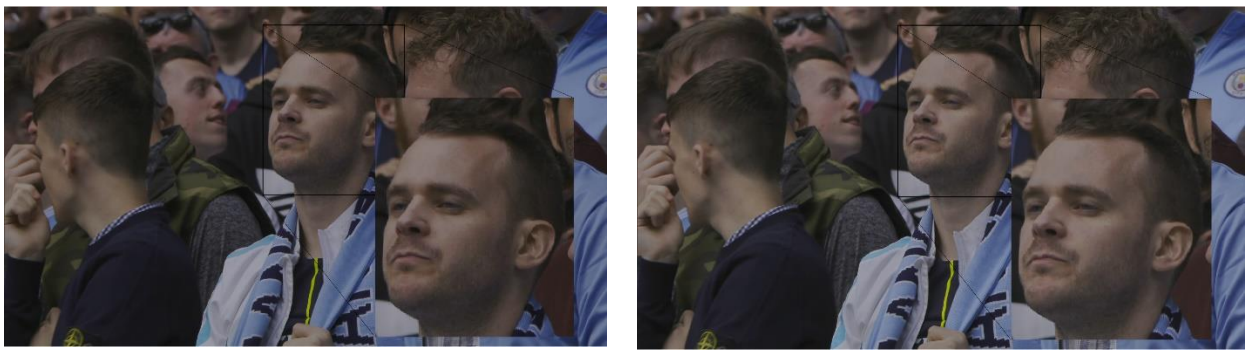
(g) Pre-scale 2 (ΔP for LG C2 (F) = 39.11W, 57.41W, 48.74W, 41.83W)

Figure 3 - Visual qualitative assessment between source, proposed JND models and the Pre-scale variants. ΔP obtained for each sequence is also presented.

Figure 4 demonstrates a visual comparison between the application of different values of ϑ and skin detection masking through Nat His and FA cup videos respectively.

(a) JND var 3 (ϑ =2.5)

(b) JND var 4 (ϑ = 1.5)



(c) JND var 1 (without skin detection mask)

(d) JND var 3 (with skin detection mask)

Figure 4 - Visual analysis of the application of skin detection and changes to luminous adaptation.

CONCLUSION

High dynamic range videos consume a significant amount of power for displaying purposes. To this end, this paper proposes a JND-based model that exploits deep learning based ROI detection, luminous adaptation and spatial correction to reduce the power consumption at minimal visual fidelity loss. Experimental results report significant power savings in OLED filmmaker modes and all modes in LCD displays. The combined analysis of power savings and visual fidelity demonstrate superior performance over the pre-scale parametric model.

Chrominance adaptations and subjective visual quality assessments will be carried out in the future. Further experiments are needed to capture the extra computational complexity that would be added to decoding devices while processing the proposed algorithms.

ACKNOWLEDGEMENTS

This work is partly funded by SEQUOIA, an Innovate UK project (grant number: 96984).

References

[1] P. Chondro, C.-H. Chang, S.-J. Ruan and C.-A. Shen, “Advanced multimedia power-saving method using a dynamic pixel dimmer on AMOLED displays,” *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 28, p. 2200–2209, 2017.

[2] J. Adhuran, G. Kulupana, C. Galkandage and A. Fernando, “Multiple quantization parameter optimization in versatile video coding for 360° videos,” *IEEE Transactions on Consumer Electronics,* vol. 66, p. 213–222, 2020.

[3] M. Dong and L. Zhong, “Chameleon: A color-adaptive web browser for mobile OLED displays,” in *Proceedings of the 9th international conference on Mobile systems, applications, and services*, 2011.

[4] C. Lajarge, F.-X. Thomas, E. Souksava, L. Chanas, H.-P. Nguyen and F. Guichard, “Objective image quality evaluation of HDR videos captured by smartphones,” *Electronic Imaging,* vol. 34, p. 1–6, 2022.

[5] Y. Sugito, J. Vazquez-Corral, T. Canham and M. Bertalmío, “Image Quality Evaluation in Professional HDR/WCG Production Questions the Need for HDR Metrics,” *IEEE Transactions on Image Processing,* vol. 31, p. 5163–5177, 2022.

[6] X. K. Yang, W. S. Ling, Z. K. Lu, E. P. Ong and S. S. Yao, “Just noticeable distortion model and its applications in video coding,” *Signal processing: Image communication,* vol. 20, p. 662–680, 2005.

[7] D. Yuan, T. Zhao, Y. Xu, H. Xue and L. Lin, “Visual JND: A perceptual measurement in video coding,” *IEEE Access,* vol. 7, p. 29014–29022, 2019.

[8] A. Liu, W. Lin, M. Paul, C. Deng and F. Zhang, “Just noticeable difference for images with decomposition model for separating edge and textured regions,” *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 20, p. 1648–1652, 2010.

[9] Q. Jiang, Z. Liu, S. Wang, F. Shao and W. Lin, “Toward top-down just noticeable difference estimation of natural images,” *IEEE Transactions on Image Processing,* vol. 31, p. 3697–3712, 2022.

[10] H. Hadizadeh, “Energy-efficient images,” *IEEE Transactions on Image Processing,* vol. 26, p. 2882–2891, 2017.

[11] J.-H. Choi, O.-Y. Lee, M.-Y. Lee, K.-J. Kang and J.-O. Kim, “JND-based power consumption reduction for OLED displays,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences,* vol. 99, p. 1691–1699, 2016.

[12] X. Chen, Y. Chen, Z. Ma and F. C. A. Fernandes, “How is energy consumed in smartphone display applications?,” in *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, 2013.

[13] M. Dong and L. Zhong, “Power modeling and optimization for OLED displays,” *IEEE Transactions on Mobile Computing,* vol. 11, p. 1587–1599, 2011.

[14] D. Kim, W. Jung and H. Cha, “Runtime power estimation of mobile AMOLED displays,” in *2013 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2013.

[15] D. J. Pagliari, S. Di Cataldo, E. Patti, A. Macii, E. Macii and M. Poncino, “Low-overhead adaptive brightness scaling for energy reduction in OLED displays,” *IEEE Transactions on Emerging Topics in Computing,* vol. 9, p. 1625–1636, 2019.

[16] A. Bhojan, “Adaptive video content manipulation for OLED display power management,” in *Proceedings of the 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2018.

[17] Y.-G. Shin, S. Park, Y.-J. Yeo, M.-J. Yoo and S.-J. Ko, “Unsupervised deep contrast enhancement with power constraint for oled displays,” *IEEE Transactions on Image Processing,* vol. 29, p. 2834–2844, 2019.

[18] H. R. Sheikh and A. C. Bovik, “A visual information fidelity approach to video quality assessment,” in *The first international workshop on video processing and quality metrics for consumer electronics*, 2005.

[19] L. Zhang, Y. Shen and H. Li, “VSI: A visual saliency-induced index for perceptual image quality assessment,” *IEEE Transactions on Image processing,* vol. 23, p. 4270–4281, 2014.

[20] P. K. Choubey, A. K. Singh, R. B. Bankapur, P. C. Vaisakh and M. Prabhu, “Content aware targeted image manipulation to reduce power consumption in OLED panels,” in *2015 Eighth International Conference on Contemporary Computing (IC3)*, 2015.

[21] J. Wu, G. Shi, W. Lin, A. Liu and F. Qi, “Just noticeable difference estimation for images with free-energy principle,” *IEEE Transactions on Multimedia,* vol. 15, p. 1705–1710, 2013.

[22] E. Sánchez-Nielsen, L. Antón-Canalís and M. Hernández-Tejera, “Hand gesture recognition for human-machine interaction,” 2004.

[23] K. He, G. Gkioxari, P. Dollár and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2017.

[24] M. Contributors, *MMSegmentation: OpenMMLab Semantic Segmentation Toolbox and Benchmark,* 2020.

[25] *FFmpeg,* 2018.