ITUPublications

International Telecommunication Union
Radiocommunication Sector

Report ITU-R BT.2540-0

(03/2024)

BT Series: Broadcasting service (television)

Display energy reduction through image signal processing

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at https://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
TF	Time signals and frequency standards emissions

***Note**: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2024

REPORT ITU-R BT.2540-0

Display energy reduction through image signal processing

(2024)

Scope

Broadcasting and streaming technologies incur a cost in terms of energy that is distributed over the entire transmission chain, from production to distribution/transmission and final viewing by consumers. Television displays, when considered the whole quantity around the world, consume a relatively large part of this energy. This energy consumption may be mitigated by content-adaptive image signal processing while minimizing the impact on visual quality. This Report describes such techniques.

Keywords

Energy reduction, television displays

Abbreviations

ABC	Automatic brightness control
BSD	Berkeley segmentation dataset
CSF	Contrast sensitivity function (see Rec. ITU-T T.800 on Information technology – JPEG 2000 image coding system: Core coding system)
CWT	Continuous wavelet transform
DCT	Discrete cosine transform
EWMA	Exponentially weighted moving average
GHG	Greenhouse gasses (see, for example, Rec. ITU-T L.1451 on Methodology for assessing the aggregated positive sector-level impacts of ICT in other sectors)
HPCCE	Histogram-based power constraint contrast enhancement
I2GEC	Image Integrity-based gray-level error control
ICT	Information and Communication Technologies (see, for example, Rec. ITU-T L.1022 on Circular economy: Definitions and concepts for material efficiency for information and communication technology)
IQPC	Image-quality-based power control
JND	Just-noticeable difference
LABS	Low-overhead adaptive brightness scaling
LAPSE	Low-overhead adaptive power saving and contrast enhancement
LCD	Liquid crystal display (see, for example, Rec. ITU-T L.1410 on Methodology for environmental life cycle assessments of information and communication technology goods, networks and services)
LPIPS	Learned perceptual image similarity
MPEG	Moving picture experts group
OLED	Organic light emitting diode
PLS	Physical layer security

PSNR Peak signal-to-noise ratio

ReLu Rectified linear unit

SJND Saliency-modulated just noticeable difference

SSIM Structural similarity index measure

1 Introduction

The consumption of energy has a direct and significant impact on our climate, as to this day a diminishing but still very large proportion of our energy needs is satisfied by the use of fossil fuels that emit greenhouse gases (GHGs) such as carbon dioxide (CO_2). Addressing climate change may proceed by reducing electricity needs, and by transitioning to sustainable sources of electricity.

An area of interest where the global use of electricity is difficult to evaluate is that of Information and Communication Technologies (ICT), of which broadcasting, streaming, and gaming form a part [1]. ICT is often seen as a sector that contributes to the reduction of energy elsewhere, but at the same time it is a significant user of electricity and therefore has a nontrivial impact on our climate.

Of all the data that is communicated over the internet, about 70% to 80% is due to video [2]. Addressing the energy efficiency of video production, communication and display therefore has the potential to have a significant influence on the total environmental impact of the ICT industry. A good proportion of the energy used to consume video content is related to the display devices and televisions [3][4]. Notably, electricity consumed to produce light in display devices has a high cost. Reduction of the amount of light produced is therefore desirable, as this helps to reduce the amount of energy necessary to operate the display. The advantage of this is two-fold: less pressure on the climate, and longer battery life in mobile devices.

Techniques currently in use in display devices to reduce their use of electricity include adaptations to the viewing environment in the form of automatic brightness control (ABC), reduction of the display brightness when a still image is displayed for a certain amount of time (for example when video is paused). Detection of the absence of viewers may be used to dim the screen as well.

Orthogonal to these measures, the content itself affords an opportunity to reduce energy consumption, as the human visual perception of spatial patterns and contrasts is often subject to masking, i.e., small changes to the content will go unnoticed due to the presence of contrast. Thus, it is possible to reduce pixel values by a small amount in such a way that a television screen uses less energy, without it affecting the visual quality of the content. Note that a small reduction of energy per device leads to a large reduction of energy globally due to the sheer number of devices in use. This leads to the first of two use cases: reduce the energy consumption of content as much as possible with the constraint that the visual quality is not or only minimally affected. A method matching this use case is presented in Annex 1.

Alternatively, a second use case is defined whereby the amount of reduction in energy use is defined first, and the best image quality is sought that satisfies this constraint. This use case would allow the user control over the amount of energy that should be saved, while maximizing visual quality. A method for this use case is presented in Annex 2.

Finally, to demonstrate the efficacy of the methods presented in Annexes 1 and 2, a set of comparisons with the state-of-the-art is provided in Annex 3 and further tests and experiments are reported in Annex 4.

2 Display energy consumption dependent on video characteristics

Energy consumption in displays depends on several factors which relate to the specification of the display and the content received and displayed. Design factors influencing the power consumption of a given display are its size, resolution, dynamic range, colour gamut, frame rate, and the specific display technology employed (currently mostly backlit liquid crystal display (LCD) screens and organic light emitting diode (OLED) screens). Some tests have found an approximately linear relationship between peak luminance and power consumption for both LCD and OLED displays, as well as a linear relationship between a display's size and power consumption[1].

The energy consumption of displays is also dependent on the content being shown. This affords an opportunity to reduce the power consumption of a display device by processing the content. This Report outlines such opportunities.

In OLED displays each pixel is independently emissive. The energy consumption of an OLED display is therefore directly proportional to the content being displayed. For LCD displays, however, the power consumption is traditionally only weakly dependent on the actual content being displayed. Nonetheless, there exist backlight scaling technologies that can reduce the power consumption in the presence of dark content. Such technologies simultaneously reduce the strength of the backlight and increase the transparency of the displayed pixels. One component of ISO/IEC 23001-11:2019 (Information technology – MPEG systems technologies – Part 11: Energy-efficient media consumption (green metadata)) enables such backlight scaling to be guided by analysing the content.

For either display type, it is possible to process the content such that, when displayed on a display, it uses less energy. For an OLED display, this processing alone will be sufficient. For LCD displays, such processing can inform a backlight scaling algorithm, and so indirectly contribute toward a lower energy expenditure. It is noted that ISO/IEC 23001-11:2019 standardises such image processing that involves either linear scaling or clamping of high luminance pixels. The present Report describes content adaptive technologies (Annex 1) that enable higher luminance reductions than could be achieved by linear scaling, as shown in the provided comparisons (Annex 2).

3 Content-adaptive energy reduction framework

A framework that enables content-adaptive energy reduction will require two main steps: an analysis step to assess the potential of each pixel/region in a frame to mask a reduction of light, and a second step to apply this reduction, possibly taking into account display parameters. It is noted that the analysis is dependent on the content only. It is further noted that such spatially varying analysis is often somewhat computationally demanding. Therefore, a suitable way to implement such a technique, is to perform the analysis for the content once, prior to broadcasting/streaming, and to adapt the content in the television, taking into account specific display capabilities and possibly the viewing environment.

A consequence is that some additional information needs to be attached to the content, in the form of metadata. Transmission of this additional information does not require significant additional energy, as energy used for transmission is significantly dependent on the available bandwidth, and only weakly dependent on the used bandwidth [5].

Thus content-adaptive energy reduction of display devices requires the following operations:

– analysis of content
– production of a map encoding the potential to reduce energy
– attachment of this pixel map to the content

---

1 https://www.rtings.com/tv/learn/led-oled-power-consumption-and-electricity-cost

– transmission of content with additional data
– display-side processing of content according to the received additional data
– final display of processed content.

The analysis, production of additional data, and the display-side processing are interlinked. The following section presents a specific method to achieve content-adaptive energy reduction.

The advantage of this framework is that the more costly processing is done once, while the adaptation of the content is done in each television set individually. The advantage for broadcasters is that this method can help curb their Scope 3 emissions[2]. The advantage for display makers is that a small adjustment to their firmware allows a reduction of energy, making it easier to conform to regulatory requirements. For consumers the advantage is that less energy is expended, the framework therefore yielding a small economic benefit as well. Finally, the planet benefits in that less energy is used, and therefore fewer green-house gas emissions are produced.

4 Content-adaptive energy reduction methods

Noting that the production of light in a television is the most energetically expensive part of a display device, reducing the amount of light produced has a direct (and linear) impact on the energy consumption of a display device. Two methods are briefly introduced in this section:

– Method A, which aims to reduce display energy consumption as much as possible without changing the visual appearance of the content.
– Method B, which allows a viewer to choose a desired reduction of display energy consumption, while minimizing the change in visual appearance.

Thus, the method presented first in this section, and defined in detail in Annex 1, allows a reduction of light that can be configured to be unnoticeable to the viewer, allowing the viewer to enjoy the content as intended by its producer. It is, however, also possible to adjust a parameter to enable stronger reduction of light, and so obtain a further reduction of energy. This parameter could be exposed to the viewer.

The method is based on the notion of a just-noticeable difference (JND), the amount a patch of light can be changed before half the observers notice the change in a direct comparison[3]. When luminance is concerned, such a JND can be obtained from a contrast sensitivity function (CSF), such as the one proposed by [6], following the process by which the PQ curve was designed [1], and adopted in Recommendation ITU-R BT.2100. Thus, for each pixel a JND is computed, and the luminance of each pixel is reduced by an amount derived from this JND. Applying the CSF to each pixel is performed within each display, as is the computation of a per-pixel JND and reducing the pixel values accordingly.

For each pixel of a given frame, Barten's model of contrast sensitivity requires as input a luminance and an angular frequency, from which the contrast sensitivity can be calculated. The input luminance is simply the luminance of each given pixel. The angular frequency, on the other hand, requires a more complex analysis. To arrive at an angular frequency for a given pixel, a wavelet analysis is appropriate, as this allows a spatially varying frequency analysis. While it is fundamentally impossible to determine exactly which frequencies are available at any given pixel location, the best trade-off between a spatial and a frequency analysis is afforded by a continuous wavelet transform

---

[2] https://www.carbontrust.com/our-work-and-impact/guides-reports-and-tools/briefing-what-are-scope-3-emissions

[3] Note that this makes the method conservative, in that in an actual use case an observer does not have access to the unprocessed image, and would therefore not be able to effect such a direct comparison.

(CWT) [8]. As the orientation of these frequencies is unimportant in this application, a CWT with an isotropic wavelet, such as the Mexican hat wavelet, is used. From this wavelet analysis, a map is constructed indicating for each pixel the frequency for which the CSF multiplied by wavelet magnitude produces the highest sensitivity.

This frequency map is then transmitted along with the content. Upon reception, a television applies the contrast sensitivity function again, and reduces pixel values according to the associated JNDs. For OLED displays, the resulting pixel values are displayed directly, leading to a reduction of energy. For backlit displays, the reduced pixel values are used in a backlight scaling approach, whereby the intensity of the backlight is reduced, and the transparency of the display panel is adjusted according to the reduced pixel values. The process is described in an implementable form in Annex 1.

For use cases where the viewer requires control over the amount of energy reduction, a method may be defined that takes as input parameter a desired energy reduction. In this case, the aim of such a method is to adjust an image such that the visual quality is maintained as much as possible, while achieving the specified energy reduction. While such a method could be developed programmatically on the basis of the method specified in Annex 1, an alternative method has been found to perform well for this use case. This method, which is based on a lightweight neural-network-based analysis of the content, is described in Annex 2. This method has a relatively small number of trainable parameters, and it is therefore efficient in execution. The network is used only for the analysis part, which in the framework described earlier in this document, would be performed prior to transmission. The content adaptation part, which would be carried out by a display device for instance, consists of simple subtraction and multiplication, and is therefore straightforward to implement in display hardware.

Annex 1

Process of content-adaptive processing to maximise energy reduction without affecting visual quality (Method A)

The purpose of the method described here is to reduce pixel values in video frames, so that a television uses less energy to reproduce the image [2]. The method is intended to provide the largest energy reduction under the constraint that the result is visually indistinct from the unprocessed image. Thus, the amount by which each pixel value is changed will be less than a given number of JNDs. Key here is that for each pixel the corresponding JND may be different. To determine the amount of reduction achievable for each pixel, JNDs are derived from a CSF, which itself takes as input the luminance of each pixel, as well as a measure of locally available frequencies obtained through wavelet analysis. The wavelet analysis produces a map of frequencies which can be transmitted along with the video content.

A receiving television then uses this map to compute a per-pixel contrast sensitivity, which in turn is used to compute a per-pixel scale factor which is to be applied to the input frame. This produces an output frame that will cause a display device to use less energy. The amount of processing can be adjusted through a free parameter which can be set so as to produce a transparent result. Alternatively, for higher values of this parameter, additional energy reduction could be achieved.

1.1 Contrast sensitivity function

Barten's contrast sensitivity function is defined as:

$$S(L,u) = \frac{M_{opt}(u)/k}{\sqrt{\frac{2}{T}\left(\frac{1}{X_0^2}+\frac{1}{X_{max}^2}+\frac{u^2}{N_{max}^2}\right)\left(\frac{1}{\eta pE}+\frac{\varphi_0}{e^{-(u/u_0)^2}}\right)}}$$

using the following quantities:

$$M_{opt}(u) = e^{-2\pi^2\sigma^2u^2}$$

$$\sigma = \frac{\sqrt{\sigma_0^2+(C_{ab}d)^2}}{60}$$

$$d = 5 - 3\tanh\left(0.4\log\left(\frac{LX_0^2}{40^2}\right)\right)$$

$$E = \frac{\pi d^2}{4}L\left(1-\left(\frac{d}{9.7}\right)^2+\left(\frac{d}{12.4}\right)^4\right)$$

In these equations, u is an angular frequency, $M_{opt}(u)$ is the optical modulation transfer function, σ models the width of $M_{opt}(u)$ which is dependent on the lens aberrations C_{ab} and pupil diameter d. The retinal illumination is modelled by E.

The constants in these equations with their values are: the signal-to-noise ratio of the eye k = 3.0, a constant modelling average visual acuity σ_0 = 0.5 arcmin, the impact of lens aberrations C_{ab} = 0.08 arcmin, the integration time of the eye T = 0.1 s, the angular extent of the object X_0 = 2°, the maximum integration angle of the eye X_{max} = 12°, the maximum number of cycles over which the eye can integrate N_{max} = 15 cycles, the quantum detection efficiency of the eye η = 0.03, the spectral density of neural noise φ_0 = 3e−8 s deg^2, the frequency above which lateral inhibition ceases u_0 = 7 cycles/deg, and the luminous flux to photon conversion factor p = 1.2e6 photons/s/deg^2/Td.

1.2 Wavelet analysis and map construction

The luminance of a pixel x is given by $L(x)$. A continuous wavelet transform C using a Mexican hat wavelet is applied to the luminance values. The wavelet pyramid consists of 10 levels i, representing spatial scales s_i of 1.00, 1.64, 2.69, 4.42, 7.26, 11.91, 19.54, 32.08, 52.64 and 86.4 pixels. The subsequent analysis is applied on the absolute values of the wavelet coefficients, which are subsequently filtered with a Gaussian convolution kernel of size $3s_i$ for coefficients at level i. This process conditions the wavelet coefficients which avoids unstable behaviour in regions that are uniform but containing noise.

For each of the spatial scales s_i a corresponding angular frequency can be constructed:

$$u_i = \frac{\pi}{180}\,\frac{0.5}{s_i}\,\frac{d\,n_x}{l_x}$$

In this equation, d is an assumed distance between the viewer and the television screen (in metres), n_x is the horizontal pixel resolution of the screen/content, and l_x is the horizontal size of the screen (in metres). An average distance of $d = 2.8\ m$ may be assumed. The horizontal distance of a 50″ television is $l_x = 1.3\ m$.

The continuous wavelet decomposition C along with the pixel luminance $L(x)$ can be used to determine the angular frequency $u(x)$ at pixel x for which the human visual system is most sensitive, given the frequencies locally available in the image. This is achieved by solving the following optimization:

$$u(x) = \underset{u_i}{\operatorname{argmax}}\, C(x, u_i)\ S(L(x), u_i)$$

As for any reasonable image size there are no more than 10 wavelet levels i necessary, the output $u(x)$ of this process has one of only 10 different values. This process is carried out for all pixels individually, so that the pixel-map of frequencies u also contains only ten different values, which are logarithmically spaced.

A filtering step is then applied to remove sharp transitions in the frequency map. As the above equation should be implemented as a loop over all wavelet levels i, while keeping track for each pixel which angular frequency produces the highest response, this filtering step can be incorporated into the same loop. Thus, a separate map $M_i(x)$ is created for each wavelet level i. Each element x in map $M_i(x)$ will have a value of either 0 or u_i.

The maps $M_i(x)$ are then filtered by Gaussian convolution with a filter kernel that depends on the angular frequency u_i associated with wavelet level i. The kernel size σ_i for wavelet level i is given by:

$$\sigma_i = \max\left(1, \min\left(512, \left(\frac{10}{u_i}\right)^4\right)\right)$$

The filtered frequency map is then constructed by summation:

$$u'(x) = \sum_i M_i \otimes G_{\sigma_i}$$

As this filtering may affect the range of values in the map, a rescaling is performed:

$$u''(x) = u'(x)\frac{\max(u)}{\max(u')}$$

This scaling is not applied when the frame is all black. Further, this scaling step may lead to temporal artifacts if applied to video. A solution to this problem is to apply leaky integration to the scaling factor max(u)/max(u'). If for frame $t-1$ the scaling factor is given by s_{t-1}, then the scaling factor s_t to be applied to frame t is given by:

$$s_t = \alpha s_{t-1} + (1-\alpha)\frac{\max(u)}{\max(u')}$$

The scaled frequency for each pixel x in frame t is then obtained by setting $u_t''(x) = u_t'(x)s_t$, with $\alpha = 0.8$ producing flicker-free results.

The map $u_t''(x)$ is attached as auxiliary data to frame t.

1.3 Content adaptation

A television receiving content will decode frame t as well as the auxiliary map $u_t''(x)$. The luminance $L(x)$ of a pixel x and its associated frequency $u_t''(x)$ is used to first determine the contrast sensitivity of this pixel by evaluating $S(x) = S\big(L(x), u_t''(x)\big)$. The contrast sensitivity value $S(x)$ is converted to a minimum detectable modulation $m(x) = 1/S(x)$. To arrive at a pixel luminance adjustment

$L_{adj}(x)$, it is observed that the minimum detectable modulation can be equated to Michelson contrast $M(x) = \big(L(x) - L_{adj}(x)\big) / \Big(L(x) + L_{adj}(x)\Big)$:

$$m(x) = M(x)$$

From this it follows that the adjustment applied to pixel x is:

$$L_{adj}(x) = L(x) \frac{1 - f\, m(x)}{1 + f\, m(x)}$$

where a free parameter f is introduced that can guide the amount of adjustment. Small values of f produce a small reduction of light (and therefore energy), while a larger value of f further reduces the demand for energy. The choice of f determines whether the resulting imagery is distinguishable from the unprocessed images or not.

The above equation shows that each pixel is multiplied by a ratio:

$$r(x) = (1 - f\, m(x)) \,/\, (1 + f\, m(x))$$

This ratio could in principle be applied to luminance only. However, the perception of luminance and chromaticness are related, and therefore an additional chrominance scaling is required. A better approach is therefore to apply the adjustment after converting the image to CIE Lab and from there to CIE LCh. Here, the L-channel represents lightness, and the appropriate reduction of lightness is therefore defined as $r(x)^{1/3}$. In addition, in CIE LCh space chroma is adjusted as:

$$C_{adj}(x) = C(x)\big(k + (1 - k)\, r(x)^{1/3}\big)$$

where $k = 0.5$ is a constant that determines the amount of chroma adjustment.

Annex 2

Process of content adaptive processing to achieve a desired amount of energy reduction (Method B)

The method presented in this annex is intended to provide the best visual quality given a desired reduction in energy. It matches the framework described in § 3 of this Report, and it is based on a neural network implementation that is both memory and energy-efficient, flexible, and lightweight.

The memory footprint and energy consumption of a neural network is strongly related to the number of trainable parameters. This number therefore needs to be as small as possible. Such a lightweight network offers the opportunity to be deployed in different environments, such as embedded hardware, video encoding or display environments.

The method [3] enables an operator to define an energy-saving rate R_e, upon which the neural network determines a pixel-wise map M_R that can be transmitted to a display device. The display device can then use this map to adjust an RGB image to achieve an energy-saving rate $R_{e,\text{final}} \leq R_e$. This is achieved by subtracting the map:

$$R_{\text{display}} = R - \frac{R_{e,\text{final}}}{R_e} M_R$$

$$G_{\text{display}} = G - \frac{G_{e,\text{final}}}{G_e} M_R$$

$$B_{\text{display}} = B - \frac{B_{e,\text{final}}}{B_e} M_R$$

Further, an alternative hue-preserving reduction of RGB values is then achieved as follows[4]:

$$R_{\text{display}} = R \frac{\left(Y - \frac{R_{e,\text{final}}}{R_e} M_R\right)}{Y}$$

$$G_{\text{display}} = G \frac{\left(Y - \frac{R_{e,\text{final}}}{R_e} M_R\right)}{Y}$$

$$B_{\text{display}} = B \frac{\left(Y - \frac{R_{e,\text{final}}}{R_e} M_R\right)}{Y}$$

In Annex 3, both approaches are evaluated (as Method B and Method B (hue preserving), respectively). In the above equations, Y represents the luma (in Yuv space) of the input image.

Thus, the display device may select an energy saving rate $R_{e,\text{final}}$ that is less than or equal to the energy saving rate R_e used to produce map M_R.

With the aid of a neural network, the map M_R can be constructed to minimise visual artefacts, and it can be directed to be as smooth as possible, which has the advantage that the map itself can be compressed prior to transmission and thereby save bandwidth.

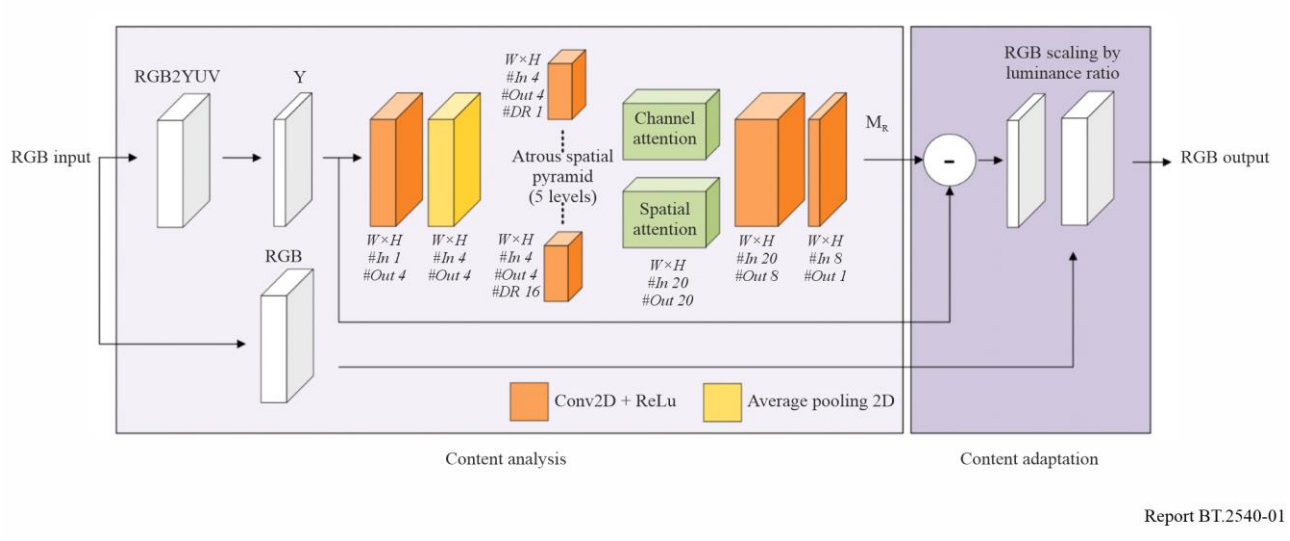
2.1 Neural network architecture

The model architecture for producing map M_R is illustrated in Fig. 1. To make the neural network as lightweight as possible, the number of trainable parameters should be kept small. This is achieved by limited the number of channels in each layer to either four or eight. Second, the method features an Atrous spatial pyramid [4] which allows the reproduction of fine-to-coarse image-level features using a small number of trainable parameters without the need for down-/up-sampling. This pyramid has five levels with dilatation rates equal to 1, 2, 4, 8 and 16. The pyramid outputs are concatenated, leading to 20 channels which are passed into channel and spatial attention layers. The channel attention layer consists of adaptive 2D average pooling followed by a rectified linear unit (ReLu) and a sigmoid activation layer. These attention layers are subsequently followed by a two ReLu-based convolution layers, which produce the final map M_R. All convolution layers use 3 by 3 kernels. The network has a total of a mere 4 832 trainable parameters.

[4] Note that this adjustment is hue-preserving because the ratio between R and G, R and B, and G and B are preserved. This method yields improved performance relative to the method presented in [3].

FIGURE 1

Model architecture



Note to Fig. 1: #In and #Out represent the number of input and output channels, respectively. #DR stands for dilatation rate.

2.2 Loss functions

To produce a network that has the aforementioned desired features, a total of four loss functions are linearly combined. Given the input and desired output luma images Y and Y_{display}, the associated display power P_Y and $P_{Y,\text{display}}$, as well as the map M_R, the loss functions are:

– The mean absolute error loss is given by $L_{\text{MAE}} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - Y_{i,\text{display}})$

– The structural similarity index measure (SSIM) loss is $L_{\text{SSIM}} = 1 - \text{SSIM}\left(Y,\ Y_{\text{display}}\right)$

– The mean per pixel power loss is defined as $L_{\text{power}} = \frac{1}{N}\left\|P_{Y,\text{display}} - P_Y(1 - R_e)\right\|^2$

– The total variance loss, defined in terms of the map M_R is given by $L_{\text{TV}} = \frac{1}{N}\sum_{i=1}^{N}((\nabla_v M_R)^2 + (\nabla_h M_R)^2)$, where ∇_v and ∇_h are the vertical and horizontal gradients.

Note that for the purpose of training, a rudimentary model of the power use of a display is given by:

$$P_Y = \frac{1}{N}\sum_{i=1}^{N} Y_i^{\gamma}$$

$$P_{Y,\text{display}} = \frac{1}{N}\sum_{i=1}^{N} Y_{i,\text{display}}^{\gamma}$$

where $\gamma = 2.2$ is the assumed gamma non-linearity employed by the display device. Note that an image is assumed to have N pixels, indexed by the variable i in the above definitions.

2.3 Training protocol

The network is trained and assessed on the Berkeley Segmentation Dataset (BSD) [5]. This dataset consists of 300 images, of which 200 for training, 40 for validation, and 60 for testing. The images have a pixel resolution of 481 by 321, in either landscape or portrait format. The images are randomly cropped into patches of size 128 by 128 pixels, which undergo random data augmentation (i.e.

horizontal flip, vertical flip and rotation of 90 degrees). The network is trained using the following parameters: Adam solver, learning rate of $1e^{-3}$, a weight decay of $1e^{-5}$, and a batch size of 4.

During the first two epochs, to ensure the quality of the output image, the loss function is only composed of the L_{MAE} and L_{SSIM} losses. The training phase converges quickly with a very good quality of reconstruction; the average peak signal-to-noise ratio (PSNR) value is above 50 dB. After these first epochs, the L_{power} and L_{TV} losses are added, to further ensure the energy reduction and the smoothness constraint on the map M_R. The full loss function is given by:

$$\alpha_{\mathrm{MAE}} L_{\mathrm{MAE}} + \alpha_{\mathrm{SSIM}} L_{\mathrm{SSIM}} + \alpha_{\mathrm{power}} L_{\mathrm{power}} + \alpha_{\mathrm{TV}} L_{\mathrm{TV}}$$

The coefficients of this linear combination are empirically set to

$\{\alpha_{\mathrm{MAE}},\ \alpha_{\mathrm{SSIM}},\ \alpha_{\mathrm{power}},\ \alpha_{\mathrm{TV}}\} = \{0.5, 1.0, 10, 0.013\}$.

Annex 3

Benchmarks and subjective evaluation

There exist several methods for reducing the energy consumption associated with specific content, of which the methods presented in this Report are part. In general, three different classes can be discerned, namely histogram-based, scaling-based, and subtraction-based. Further classifications can be made based on whether the processing is linear or non-linear and whether the method is based on a (deep) neural network or not. In addition, some methods aim to preserve quality while reducing power requirements as much as possible, while others fix a power reduction requirement and within this constraint aim to maintain visual appearance. Of the methods currently known (and summarized in Table 1), twelve were chosen for a benchmark. These are listed in **bold** in the Table, forming a representative selection of methods.

TABLE 1

Taxonomy of energy-aware image processing methods

Model	Type	Spatially varying	Linear/Non-Linear/Deep	Quality/ Power	Summary
I2GEC [8]	Hist	–	NL	Q	Hard thresholding
HPCCE [9]	Hist	–	NL	Q, P	Contrast enhancement
IQPC [10]	Hist	–	NL	Q	Roll-off clipping
PQPR [11]	Scaling	–	L	Q	SSIM-based scaling
S-VS [12]	Scaling	Yes	NL	–	Helmholtz-Kohlrausch effect
SJND [13]	Scaling	Yes	NL	Q	JND and saliency-based scaling
LAPSE [14]	Scaling	–	NL	Q, P	Polynomial scaling
LABS [15]	Scaling	–	L	Q, P	Linear scaling
PLS [16]	Scaling	–	L	P	Power-rate reduction scaling

Model	Type	Spatially varying	Linear/Non-Linear/Deep	Quality/ Power	Summary
ACE Net [16]	Scaling	Yes	D	Q, P	CNN-based contrast enhancement
DeepBattery [17]	Scaling	Yes	D	Q, P	Encoder-decoder network

TABLE 1 (*end*)

Model	Type	Spatially varying	Linear/Non-Linear/Deep	Quality/ Power	Summary
EWMA [18]	Scaling	–	NL	–	Content-dependent filter
R-ACE net [19]	Subtract	Yes	D	Q, P	CNN-based attenuation map
Method A [2]	Scaling	Yes	NL	Q	CSF-based scaling
Method B [3]	Subtract	Yes	D	Q, P	CNN-based attenuation map

Among the histogram-based methods, I2GEC and IQPC search a clipping point at which the luminance of the content is clipped. This is done while remaining above a target PSNR value. The HPCCE method performs histogram equalization guided by an energy consumption model. This jointly reduces power consumption and increases global contrast. HPCCE is therefore not transparent in terms of visual quality.

Scaling-based methods modify all pixels in the image by applying a scaling factor to each. The scaling factor can be the same or it can be different for each pixel. The following methods apply the same scaling factor to all pixels. The exponentially weighted moving average (EWMA) method recursively applies a gain-offset model to the luma component in YC_bC_r color space. The physical layer security (PLS) method reduces luma with a scaling factor k determined by the desired percentage power reduction R using $k = \left(1 - \frac{R}{100}\right)^{1/\gamma}$. The PQPR method instead determines its scaling factor k on the basis of a target SSIM value. Each of these methods either consider quality or energy reduction targets, but not both jointly.

The LAPSE and LABS methods, on the other hand, do incorporate a trade-off between quality and power reduction. LAPSE makes use of a third order polynomial to reduce luma. The LABS method bases its scaling factor on both SSIM and a desired power reduction.

The following scaling-based methods are spatially variant. The SJND method aims to preserve visual quality through the calculation of a just-noticeable difference obtained through a saliency model, in link with a contrast sensitivity function and an analysis of spatial frequencies in the discrete cosine transform (DCT) domain. Method A also falls in this class, and it is described in detail in Annex 1 of this Report. The S-VS method exploits the Helmholtz-Kohlrausch effect, in which the perception of a pixel's luminance is modified by its level of saturation.

Finally, several models are known which uses deep learning to reduce the power consumption of content without unduly affecting visual quality. Typically a network is designed along with loss functions that drive the desired power reduction and other loss functions that aim to maintain image quality. The ACE network is designed to directly produce a power-reduced image. With a different architected, the R-ACE network outputs an attenuation map which is later added to the input image to produce a power-reduced image. Finally, Method B, as described in Annex 2 of this Report is a further development of R-ACE, with a significantly simplified architecture while improving on its performance.

3.1 Benchmark with objective metrics

The methods are benchmarked using two objective metrics on images drawn from two datasets. The datasets are the Kodak24 set [20] and BSDS300 test set [21]. The evaluation uses all 24 images from the Kodak24 set and 60 randomly drawn images from the BSDS300 database. The objective metrics used for the evaluation are the SSIM [7] and the learned perceptual image similarity (LPIPS) [22]. SSIM is a perception-based method that measures image degradation between an input image and a processed image as the perceived change in structural information [7]. The values obtained with SSIM range between 0 and 1, where values close to 1 mean that the difference between the input and the processed image is small. High values are therefore better. LPIPS is a neural network-based image similarity metric [22]. The values obtained with this network also range from 0 to 1, but in this case lower values means greater similarity between input and processed images.

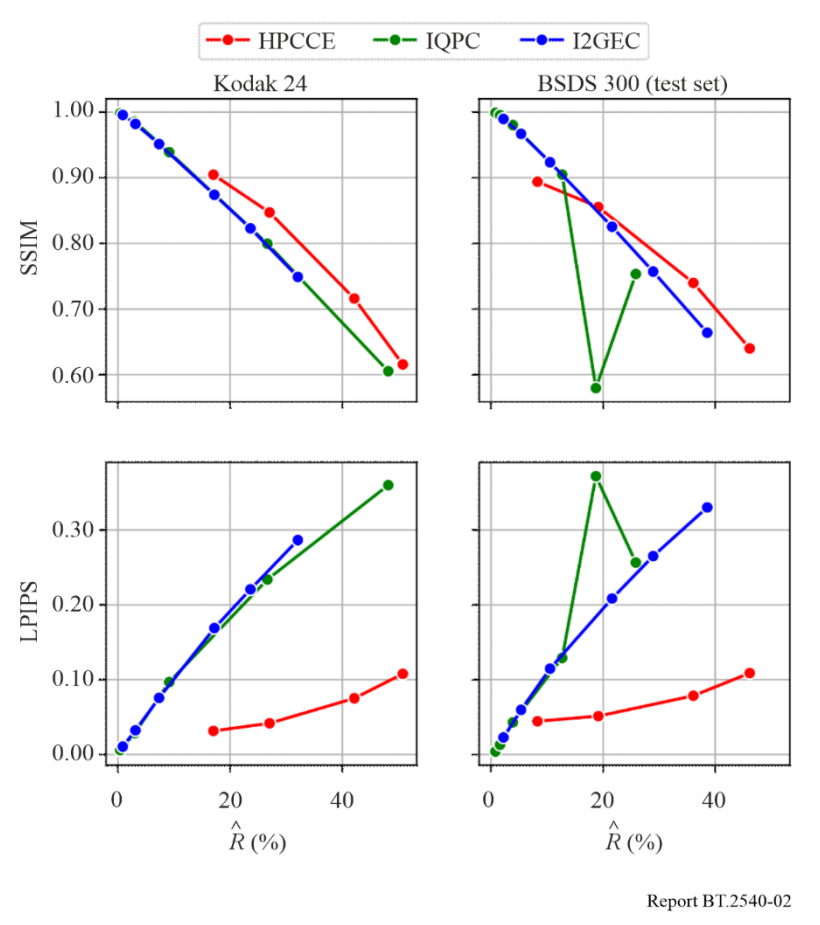
To understand the power consumption associated with each image for a specific OLED display, which has red, green, blue and white sub-pixels, a power model was developed [23]. For a given RGB pixel $(r_{\text{in}}, g_{\text{in}}, b_{\text{in}})^T$ as input, the model begins by determining red, green, blue and white values for the four sub-pixels. This is achieved by using Murdoch's model [24] which assumes that the display gamma γ is known, and that a rotation matrix R_{rot} has been derived which transforms the input RGB values into the RGB color space corresponding to the red, green and blue sub-pixels. It further assumes that the white point $(r_w, g_w, b_w)^T$ of the display is available. The (r, g, b, w) sub-pixels are then computed as follows:

$$\begin{pmatrix} r' \\ g' \\ b' \end{pmatrix} = R_{\text{rot}} \begin{pmatrix} r_{\text{in}} \\ g_{\text{in}} \\ b_{\text{in}} \end{pmatrix}^{\gamma} \begin{pmatrix} 1/r_w \\ 1/g_w \\ 1/b_w \end{pmatrix}^{T}$$

$$c_{\min} = \min(r', g', b')$$

FIGURE 2

Objective evaluation of histogram-based methods (HPCCE, IQPC and I2GEC) for different power reduction values



Note to Fig. 2: The plots at the top show SSIM values, while the bottom plots show LPIPS values.

$$\begin{pmatrix} r \\ g \\ b \\ w \end{pmatrix} = \begin{pmatrix} r' - c_{\min} \\ g' - c_{\min} \\ b' - c_{\min} \\ c_{\min} \end{pmatrix} \begin{pmatrix} r_w \\ g_w \\ b_w \\ 1 \end{pmatrix}^{T}$$

Assuming that (r, g, b, w) are normalised, a pixel that is supplied with these values will then consume an amount of power that can be estimated as:

$$P_{r,g,b,w} = r\, P_{R,\max} + g\, P_{G,\max} + b\, P_{B,\max} + w\, P_{w,\max}$$

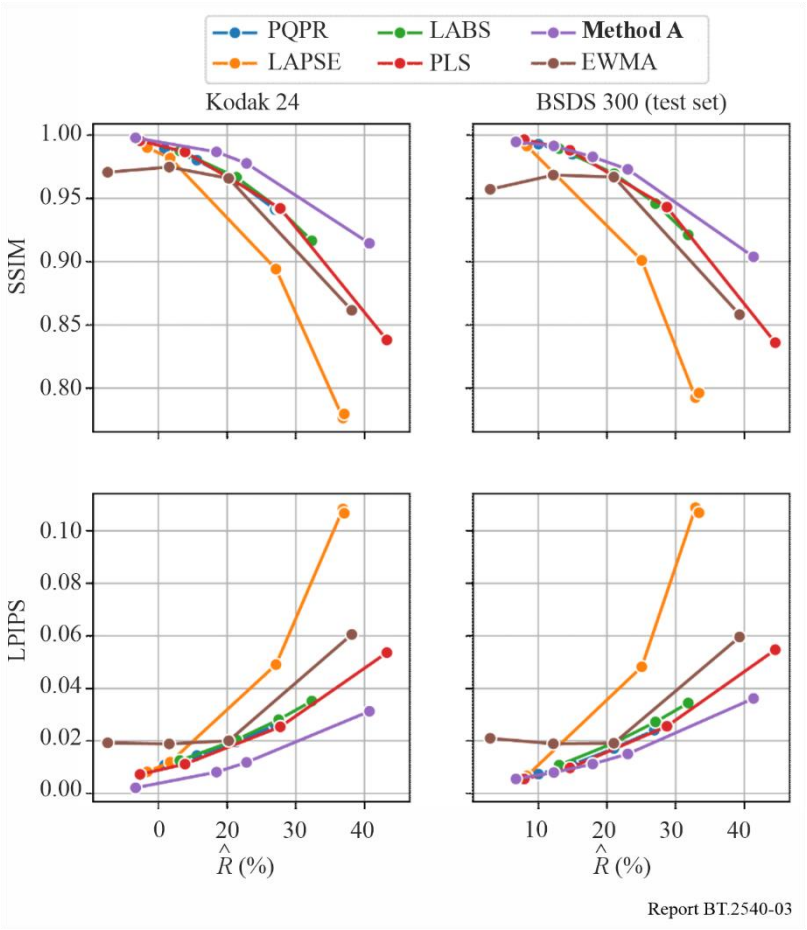
The power consumption P_{im} associated with an image is then the sum of the power consumption of each pixel. This model can be used to evaluate the power consumption associated with an image.

Any method that reduces the power consumption of a display when it is displaying said image, can be evaluated in terms of this method. The input and output images (I_{in} and I_{out}) will consume an estimated amount of power P_{in} and P_{out}, respectively. The power reduction $\hat{R}$ can then be expressed as a percentage:

$$\hat{R} = 100\left(1 - \frac{P_{\text{out}}}{P_{\text{in}}}\right)$$

FIGURE 3

Objective evaluation of scaling-based methods (PQPR, LAPSE, LABS, PLS, EWMA and Method A) for different power reduction values



Note to Fig. 3: The plots at the top show SSIM values, while the bottom plots show LPIPS values.

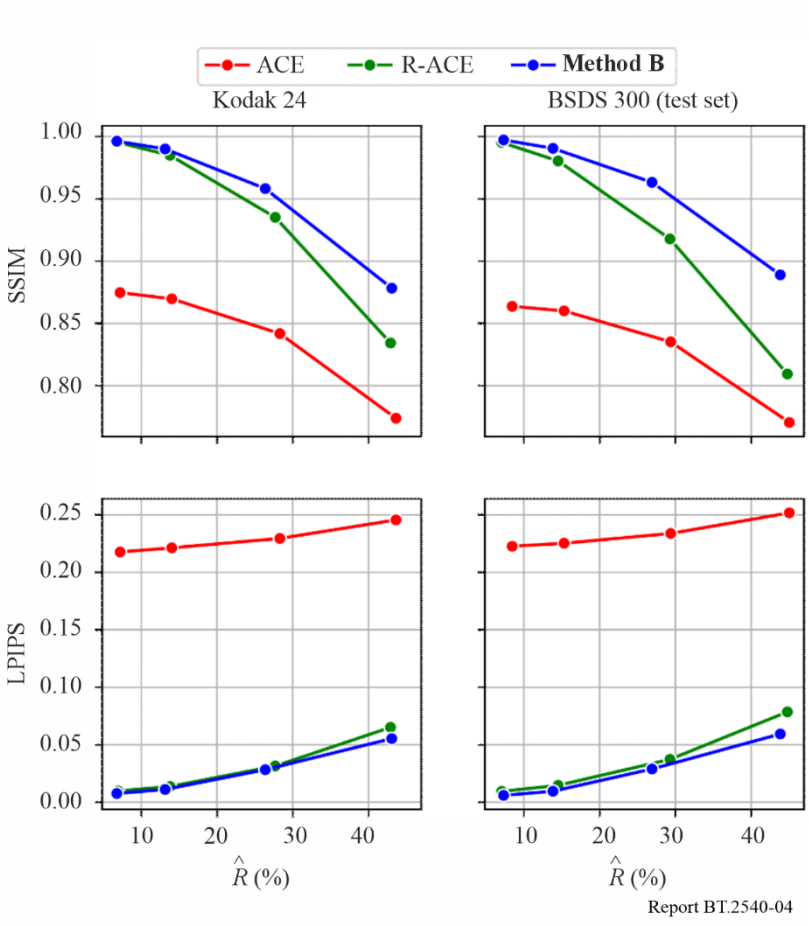
For each energy-aware image processing method, their parameters were varied to cover a range of power reductions. Methods based on histogram clipping were configured for PSNR values of 20, 25, 30 and 35. For the scaling-based models PQPR and LAPSE, the SSIM was used to determine parameter settings. Here, SSIM values of 0.88, 0.90, 0.95 and 0.99 were used. For the PLS method the luminance scaling was calculated for target power reductions of $\hat{R} = 10$, 20, 40 and 60. The α parameter of the LABS method was varied in the range between 0.5 and 2. In each case the parameters were chosen to be reasonably close to target power reduction values of $\hat{R} = 10$, 20, 40 and 60.

Figure 2 shows the results obtained for the histogram-based methods involved in the comparison. The top two plots show SSIM values as function of obtained power reduction $\hat{R}$. The bottom two results show LPIPS values as function of obtained power reduction $\hat{R}$. The plots on the left are for the Kodak24 set, whereas the plots on the right are for the BSDS300 dataset. The results for scaling based methods are shown in Fig. 3, and the results for the neural-network-based methods are shown in Fig. 4.

In general, a stronger power reduction produces a lower SSIM value and a higher LPIPS value, which is consistent with expectation. For the histogram-based methods, IQPC and I2GEC perform similarly, although it should be noted that in these evaluations the IQPC method has produced an outlier (see Fig. 2). For this class of methods HPCCE offers the best trade-off between quality and power reduction. Within the class of scaling-based methods, the best trade-off is achieved by Method A, as described in this Report, as shown in Fig. 3. Between the neural-network-based methods, it is Method B that produces the best trade-off between quality and power reduction, as can be seen in Fig. 4.

FIGURE 4

Objective evaluation of neural-network-based methods (ACE, R-ACE and Method B) for different power reduction values



Note to Fig. 4: The plots at the top show SSIM values, while the bottom plots show LPIPS values.

Figure 5 compares the best performing methods from each of the three classes (HPCCE, Method A and Method B). This plot shows that in terms of SSIM, both Method A and Method B perform similarly, and outperform HPCCE. The LPIP metric indicates an advantage for Method A relative to Method B, while both Methods A and B outperform HPCCE.

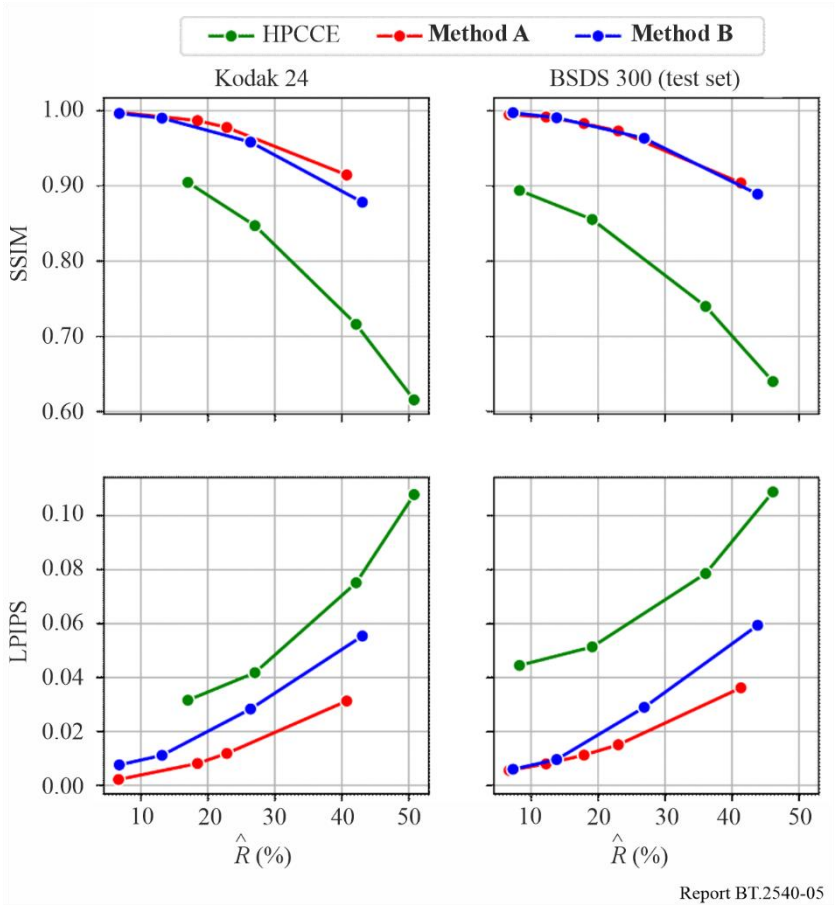
Further, to understand the effect of producing a hue-preserving power reduction, Method B was further evaluated with and without a hue preserving reduction, as presented in Annex 2. Both variants are also compared with Method A (which has its own colour management approach), as well as with the PLS method with and without hue-reserving power reduction. The results produced for the images from the Kodak24 set are shown in Fig. 6. As can be seen, both PLS and Method B improve significantly when a hue-preserving reduction is introduced. In the context of the SSIM metric, Method B now produces the best results, whereas for the LPIPS metric, Method A and PLS perform best.

3.2 Subjective test

Given the results of the benchmark presented in § 3.1, the best performing methods were subjected to an additional psychophysical evaluation. The methods in question are PLS (hue preserving), Method A and Method B (hue preserving), and for each method for different levels of energy reduction were created.

FIGURE 5

Plots comparing the best methods from each of the three classes, namely histogram-based (HPCCE), scaling-based (Method A) and neural-network-based (Method B)



The protocol involved pairwise judgements on a Likert-like scale, where similarity between two images presented side-by-side was judged. The scale ranged from ‘very dissimilar’, ‘dissimilar’, ‘neutral’, ‘similar’, to ‘very similar’. Each pair of images was presented side-by-side for 5 seconds, and a decision had to be recorded within those 5 seconds by checking one of five checkboxes. The first ten images were discarded from analysis, as these are intended for training the participant. These ten images were followed by 250 trials on images extracted from the Kodak 24 dataset. The experiment was carried out by forty participants (28 male, 12 female).

The experiment was carried out using an LG OLED display in a controlled viewing environment (dark walls and the only illumination in the room came from the display itself).

The average subjective scores obtained with this protocol are shown in Fig. 7. The lines indicate mean subjective scores, whereas the grey zone around each line represents the standard error. These results generally show a correlation between reducing the energy consumption and the detectability of differences with the input image. The PLS method (hue preserving) scores lower than Method A and Method B (hue preserving), whereas the latter two methods perform equivalently in this experiment.

FIGURE 6

Plots comparing hue preserving variants for PLS and Method B with the original versions of PLS and Method B as well as with Method A

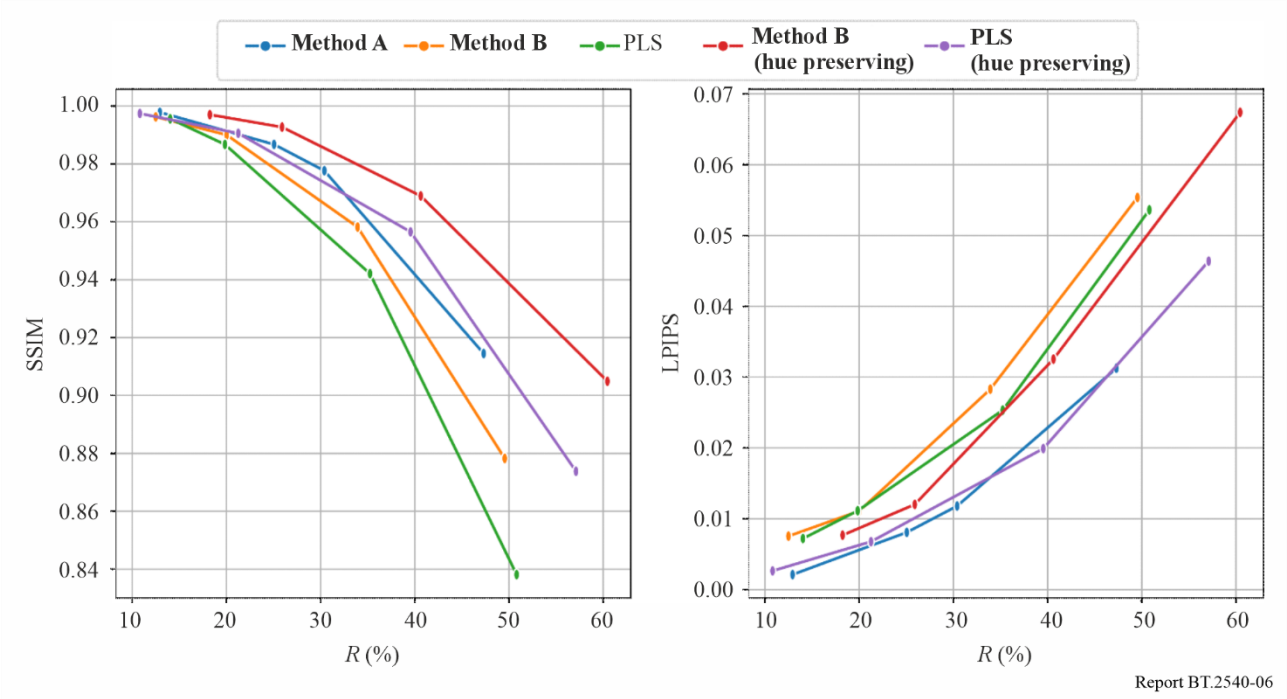
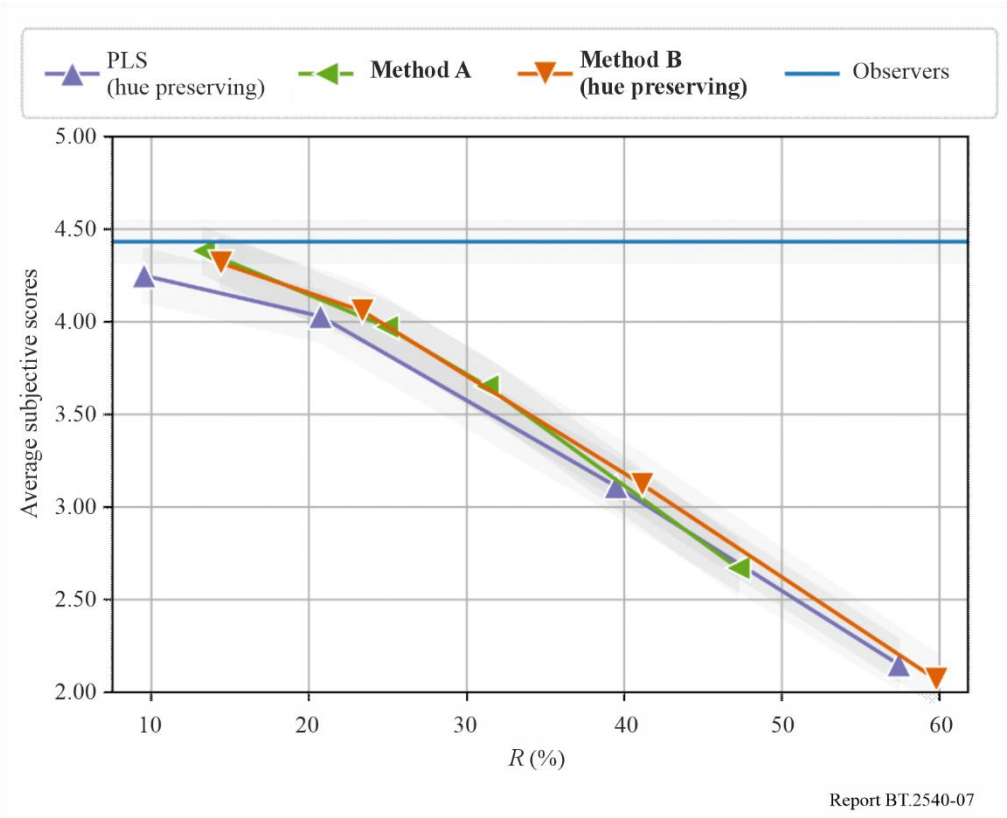


FIGURE 7

Mean subjective scores for PLS (hue preserving), Method A, and Method B (hue preserving)



Note to Fig. 7: The blue 'Observers' line represents the score given when the input image was judged against itself. The grey zones around each line indicate the standard error.

The horizontal blue 'Observers' line reports the score given to comparisons of input images with itself. As this line is not at 5.0, this means that even when the input image is shown twice side-by-side, some participants on occasion report small differences. It is interesting to note that for Method A and Method B (hue preserving) the standard error zones overlap with the standard error zone of the horizontal 'Observers' line for power reductions up to around 20%. This means that with either method power reductions up to 20% produce to imagery that, when placed side-by-side, are difficult to distinguish from the input imagery. This goes toward the suggestion that a 20% power reduction can be achieved with either method.

Annex 4

Further experiments

First, the process for achieving energy reduction in display devices by means of adjusting the content, as described in Annex 1, is evaluated in this section. This process is labelled A in this Annex. The neural network-based method described in Annex 2 is also evaluated, and it is labelled B in this Annex.

Derived and alternative methods are introduced as follows:

– Method C: This method is the same as Method A, but removes the spatially varying wavelet analysis, and uses a fixed frequency estimate instead. This method is included to test the relative merit of the wavelet analysis of Method A.

– Method D: This method implements linear scaling, as this is the most basic way in which the display power consumption associated with an image can be reduced.

4.1 Performance of Method A

To assess the performance of Method A, images are created for different values of free parameter f, ranging between 5 and 75 in increments of 5. This parameter is described in § 1.3 of Annex 1.

Method A is also compared against two alternatives. The first alternative, labelled C, omits the spatial analysis step. Barten's contrast sensitivity function is seeded with a fixed angular frequency of 5 cycles per degree, instead of a per-pixel angular frequency:

$$S_{\text{fixed}}(x) = S(L(x), 5)$$

The contrast sensitivity $S_{\text{fixed}}(x)$ is then used to compute an adjusted image $L_B(x; f)$ as in Method A. The angular frequency of 5 cycles per degree was chosen because human vision (and Barten's model) is most sensitive at this frequency. This, therefore, produces an alternative that is conservative which is desirable in a broadcast scenario where preserving visual quality is paramount. This variant could be implemented in its entirety in a display device without requiring spatial processing.

The second alternative is a simple linear scaling, labelled D. To determine the scaling factor that would produce a comparable result, for each image the mean ratio r between input image and the result obtained with Method A is determined:

$$r = \frac{1}{n_x\, n_y} \sum_x \frac{L_{\text{A}}(x;f)}{L(x)}$$

The ratio is then applied to the input image to produce a linearly scaled image $L_C(x; f)$ that has on average the same reduction of values as that achieved by Method A:

$$L_C(x; f) = r\, L(x)$$

For the purpose of comparison, a dataset of 157 images known as the INRIA Holidays database was used [6]. Method A has a single free parameter f, which can be varied to drive the strength of the processing. In the following, parameter f is varied between 5 and 75, in increments of 5, leading to fifteen different results per input image for each of Methods A, C and D.

During a critical viewing session, for each of the three algorithmic variants, the largest value of f was determined for which no differences were visible in a direct side-by-side comparison with the input image. The viewing environment is a room painted matte black, has no windows and no light other than that emanating from the screen used for the experiment. This screen showed images side-by-side on a grey background, with the input image on the left and a processed image on the right. For each image and for each variant, a trained expert was able to go back and forth between different values of f. Using this protocol, the expert selected the value of f for which no differences with the input image were visible.

In the following, these threshold values are denoted $f_{\text{threshold}}$.

To understand whether differences between the three methods in threshold values also lead to differences in light reduction, for each image the corresponding reduction of light (in cd/m^2, assuming a display with 100 cd/m^2 peak luminance) was computed as follows:

$$\Delta L_{\text{A}}(x; f_{\text{threshold}}) = L(x) - L_{\text{A}}(x; f_{\text{threshold}})$$

Analogously, light reductions for Methods C and D, labelled $\Delta L_{\text{C}}(x; f_{\text{threshold}})$ and $\Delta L_{\text{D}}(x; f_{\text{threshold}})$, are computed, the results of which are shown in Fig. 8. This Figure shows the merit of Method A, in that on average this method produces a more significant light reduction than Methods C and D.

FIGURE 8

A histogram showing the number of images for each reduction in luminance for Methods A, C and D

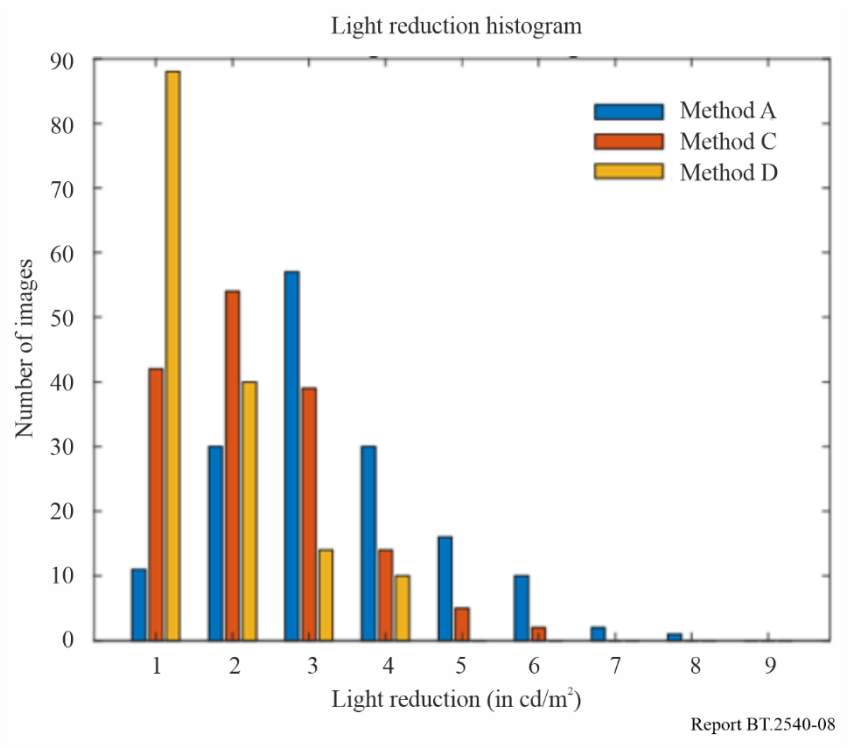
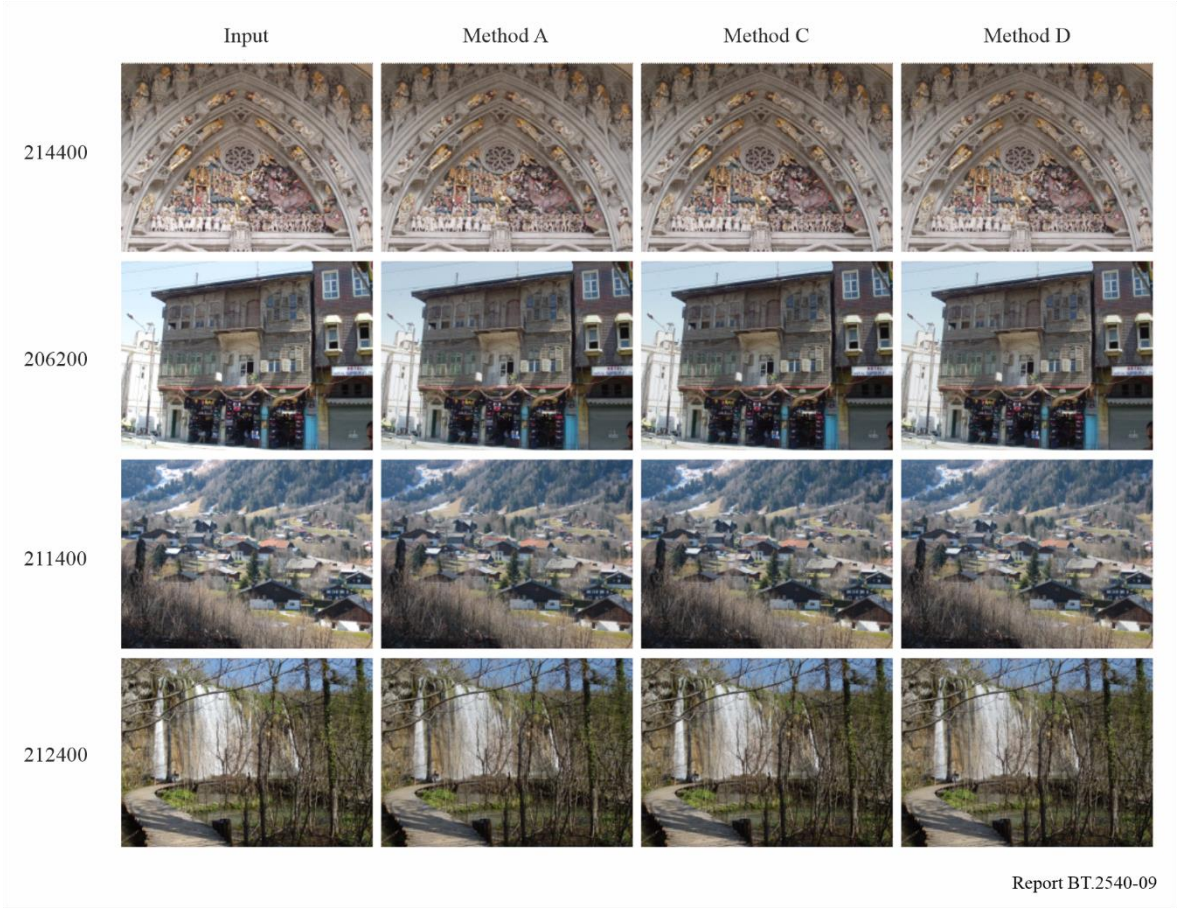


FIGURE 9

Some representative results from the INRIA Holidays database



Note to Fig. 9: The numbers on the left are the file names of the images shown.

To assess visual quality, the image selected by the expert viewer were subjected to analysis with the PSNR and the SSIM [7]. It is noted, however, that the PSNR metric is only included for completeness, as this metric is sensitive to peak values in the images, and this is what the algorithm deliberately reduces – in this use case it is therefore more a measure of success than an image quality metric. The SSIM metric combines three types of measurement to assess image differences, namely luminance, contrast, and structure. It is applied in a windowed fashion on pairs of images. A pair of identical images would produce a value of 1, where is two images that are maximally different would produce a value of 0.

The results are shown in Table 2, indicating that on average the images selected for Method A have a somewhat lower PSNR than Methods C and D, suggesting that the peak luminance has reduced a little more. The SSIM results, however, are not significantly different among the three variants.

Figures 9 and 10 present several example results as well as numerical results, as obtained by the expert viewer. Here, 3D plots show difference images $\Delta L_{\mathrm{A}}(x; f_{\mathrm{threshold}})$, $\Delta L_{\mathrm{C}}(x; f_{\mathrm{threshold}})$ and $\Delta L_{\mathrm{D}}(x; f_{\mathrm{threshold}})$ as height fields. The red line in the backplanes of the plots shows the mean luminance reduction, whereas the blue line shows the maximum reduction. The spatial variation in light reduction is significantly more pronounced in Method A compared with Method C, allowing an increased overall reduction of image luminance.

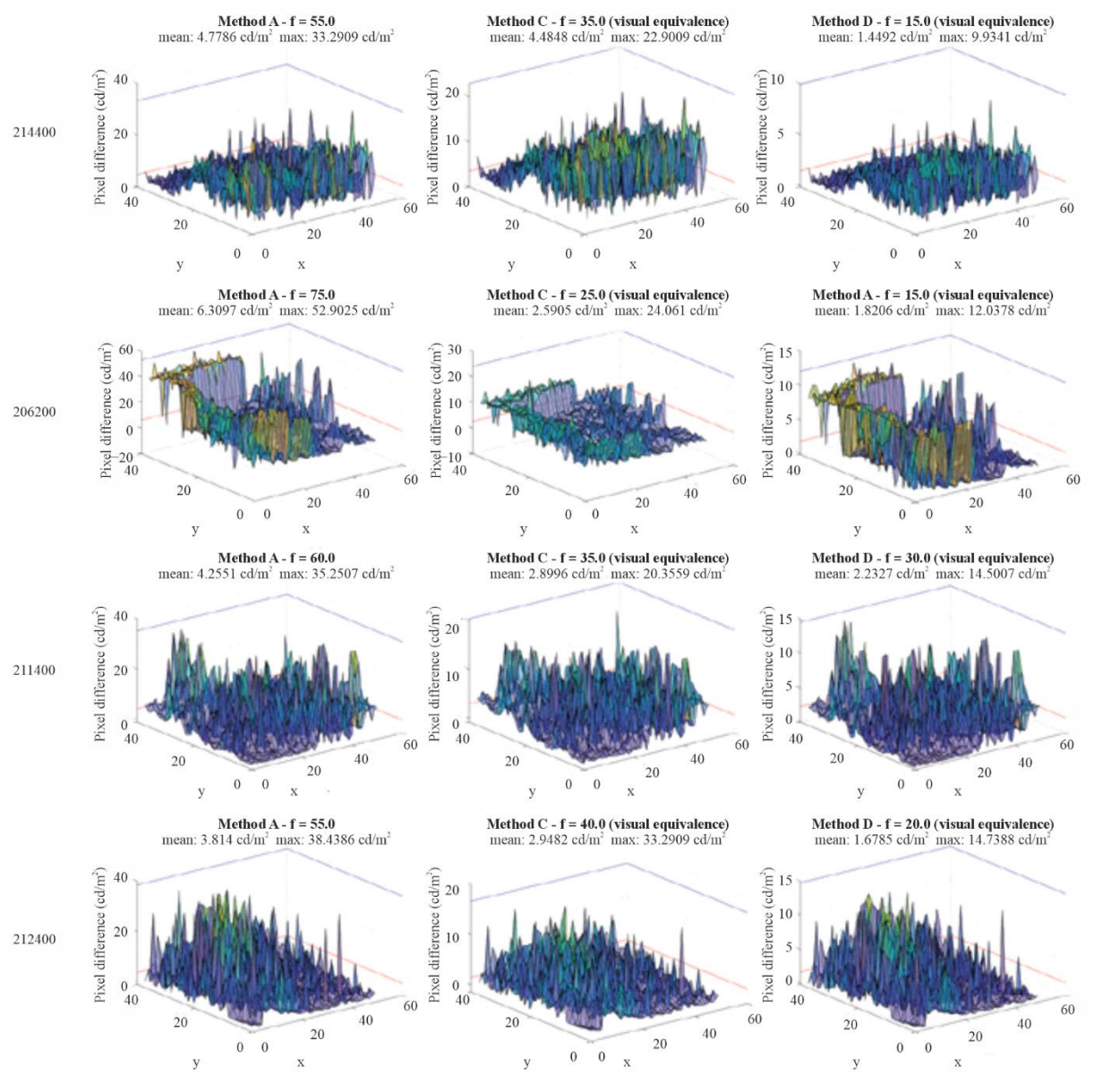
TABLE 2

Image quality metrics applied to the images selected by the expert viewer (157 images per method)

Metric	Method A	Method C	Method D
PSNR (dB)	29.67	33.25	37.52
SSIM	0.98	0.98	0.99

FIGURE 10

For each example of Fig. 9, the plots show height fields for light reduction achieved with Methods A, C and D relative to the input images



The average reduction of light output, assuming a 100 cd/m^2 peak display luminance, over all 157 images of the INRIA Holidays dataset, amounts to 3.36 cd/m^2 for Method A, 2.25 cd/m^2 for Method C and 1.55 cd/m^2 for simple linear scaling (Method D). Thus, according to this relatively limited test, Method A is able to reduce light output twice as much as could be achieved with linear scaling.

It should be noted that these results are obtained by an expert viewer in a dark room, and in a side-by-side comparison. In a real-world scenario, viewers would not have access to the unprocessed imagery, and this would allow a significantly larger reduction in luminance before the content would be seen as noticeably reduced in luminance.

Further, the results obtained with Method C show that the wavelet-based analysis used in Method A significantly adds to the ability to reduce luminance, even if the light reduction is already improved relative to linear scaling (Method D).

4.2 HDR Tests (Methods A, C and D)

Methods A, C and D are also evaluated on HDR images, which are taken from Table 8 of Report ITU-R BT.2245-9 – HDTV and UHDTV including HDR-TV test materials for assessment of picture quality. For these images, results are computed for a value of $f = 75$. For a set of 104 images, the average reduction is 13.7 cd/m^2 for Method A, 11.3 cd/m^2 for Method C and 13.3 cd/m^2 for Method D. These values are more or less similar for equivalent parameter settings. However, the distribution of this reduction over the pixels will be different, and therefore for the same parameter settings, a difference in visual quality is to be expected.

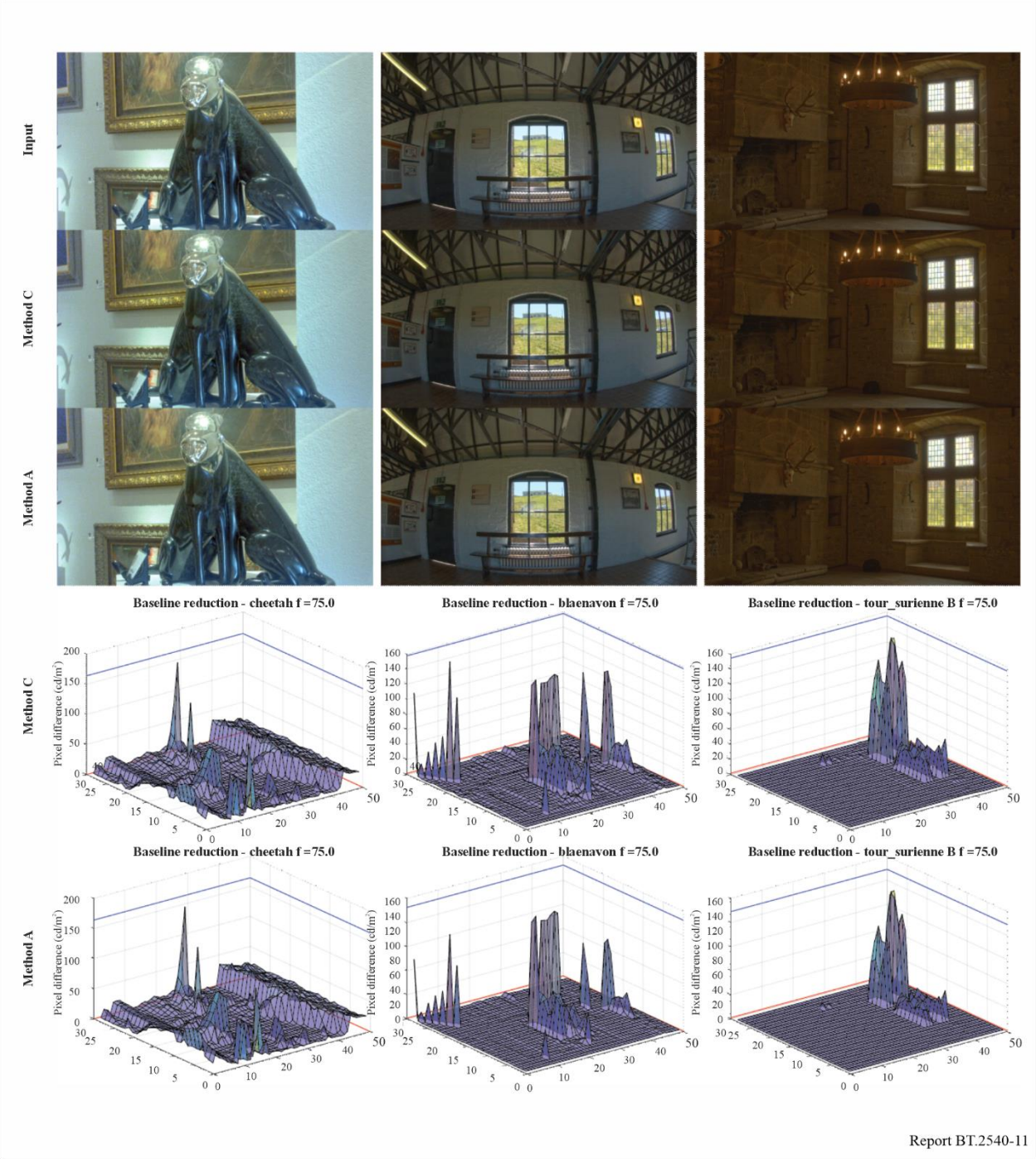
The change in peak luminance is assessed as well. Here, over the same set of images, the peak luminance is reduced by on average 180 cd/m^2 for Method A, 156 cd/m^2 for Method C, and 201 cd/m^2 for Method D.

It is noted here that for equivalent parameter settings, Method D reduces the peak luminance more than Method A. Combined with the observation that the mean luminance reduction is similar for Methods A and D, this means that the distribution of luminance reduction over the image is different for the two methods, resulting in a higher final peak brightness for Method A.

Some example results are shown in Fig. 11.

FIGURE 11

Example HDR results



4.3 Metadata tests (Method A)

Given that the frequency map produced by Method A is intended to be transmitted as metadata, the question arises as to what happens if the metadata goes missing, or arrives corrupt. To test the case of corrupt data, a simulation was performed whereby a block (a quarter the size of the frequency map) in the centre of the frequency map is randomly permuted, thus producing noise. Images with values of *f* as chosen by the expert viewer were produced in this manner.

For this test, the differences between images with and without corrupted metadata are computed. The mean of the absolute value of the differences found over all 157 images is 1.4 codeword values (i.e. on a range of 0 to 255). Example difference images are shown in Fig. 12, where 0 difference is

mapped to middle grey, darker and lighter values represent negative and positive differences. Note that both the mean and maximum differences are very small, indicating that corrupted metadata has a minimal effect on the quality of the final result.

A second simulation was carried out by replacing the centre block in the frequency map with zeros, to test the case where a part of the metadata is missing. Given that the analysis step of Method A does not ordinarily create zero values in the map, the absence of metadata in (parts of) the map can be detected. If all metadata is missing, then the method replaces the map with a fixed frequency applied to all pixels. If only some of the metadata is missing, the missing values are set to the average of the data that is present. An example of the latter case is shown in Fig. 13 (right). Note that this produces a result that is essentially indistinguishable from the case whereby metadata is preserved (i.e. the part of the image around the central block).

For this test, difference images were also computed, showing much the same behaviour as the case where metadata is corrupted. The mean of the absolute values of all differences computed over all 157 images is 1.17 codeword values. Example results of image differences are shown in Fig. 12. A difference of 0 is mapped to middle grey, whereas positive and negative differences are lighter and darker.

FIGURE 12

Example results for corrupted metadata

Report BT.2540-12

Note to Fig. 12: Shown here is a processed image (left), a processed image with frequency map randomly permuted (i.e. corrupted; middle), and the difference map between the two images (right).

FIGURE 13

Example results for missing metadata

Report BT.2540-13

Note to Fig. 13: Shown here is a processed image with the middle section of the frequency map set to zero (i.e. missing; left), and the difference map between the left image and the input image (right).

References

[1] S. Miller, M. Nezamabadi and S. Daly, “Perceptual Signal Coding for More Efficient Usage of Bit Codes”, SMPTE Motion Imaging Journal, vol. 122, no. 4, pp. 52-59, 2013.

[2] E. Reinhard, C.-H. Demarty and L. Blondé, “Pixel Value Adjustment to Reduce the Energy Requirements of Display Devices”, SMPTE Motion Imaging Journal, vol. 132, no. 7, 2023.

[3] O. Le Meur, C.-H. Demarty and L. Blondé, “Deep Learning-Based Energy Aware Images”, in IEEE International Conference on Image Processing, Kuala Lumpur, 2023.

[4] L.-C. Chen, G. Papandreou, F. Schroff and A. Hartwig, "Rethinking Atrous Convolution for Semantic Image Segmentation", arXiv 1706.05587, 2017.

[5] D. Martin, C. Fowlkes, D. Tal and J. Malik, "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics", in Proceedings of the 8th International Conference on Computer Vision, 2001.

[6] H. Jegou, M. Douze and C. Schmid, "Hamming embedding and weak geometry consistency for large scale image search", in Proceedings of the 10th European Conference on Computer Vision, 2008.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment from error visibility to structural similarity", IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.

[8] S.-J. Kang and Y. H. Kim, "Image integrity-based gray-level error control for low power liquid crystal displays", IEEE Transactions on Consumer Electronics, 55(4):2401-2406, 2009.

[9] C. Lee, C. Lee, Y.-Y. Lee and C.-S. Kim, "Power-constraint contrast enhancement for emissive displays based on histogram equalization", IEEE Transactions on Image Processing, 21(1):80-93, 2012.

[10] S.-J. Kang, "Image-quality-based power control technique for organic light emitting diode displays", Journal of Display Technology, 11(1):104-109, 2015.

[11] S.-J. Kang, "Perceptual quality-aware power reduction technique for organic light emitting diodes", Journal of Display Technology, 12(6):519-525, 2016.

[12] T. Shiga and S. Kitahara, "Power reduction of OLED displays by tone mapping based on Helmholtz-Kohlrausch effect", IEICE Transactions on Electronics, 100(11):1026-1030, 2017.

[13] H. Hadizadeh, "Energy-efficient images", IEEE Transactions on Image Processing, 26(6):2882-2891, 2017.

[14] D. J. Pagliari, E. Macii and M. Poncino, "LAPSE: Low-overhead adaptive power saving and contrast enhancement for OLEDS", IEEE Transactions on Image Processing, 27(9):4623-4737, 2018.

[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, 9(3):1625-1636, 2021.

[16] Y.-G. Shin, S. Park, M.-J. Yoo and S.-J. Ko, "Unsupervised deep power saving and contrast enhancement for OLED displays", arXiv preprint arXiv:1905.05915, 2019.

[17] J.-L. Yin, B.-H. Chen, Y.-T. Peng and C.-C. Tsai, "Deep battery saver: End-to-end learning for power constraint enhancement", IEEE Transactions on Multimedia, 23:1049-1059, 2020.

[18] M. Trigka, E. Dritsas and K. Moustakas, "Joint power and contrast shrinking in RGB images with exponential smoothing", In 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), 1-5, 2022.

[19] K. A. Nugroho and S.-J. Ruan, "R-ACE network for OLED image power saving", in 2022 IEEE 4th Global Conference on Life Sciences and Technologies (LifeTech), 284-285, 2022.

[20] R. Franzen, "Kodak lossless true color image suite", https://r0k.us/graphics/kodak/, last accessed 27/11/2023.

[21] D. Martin, C. Fowlkes, D. Tal and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics", In Proceedings of the 8th International Conference on Computer Vision, volume 2, 416-423, 2001.

[22] R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric", CoRR, abs/1801.03924, 2018.

[23] C.-H. Demarty, O. le Meur and L. Blondé, "Display power modelling for energy consumption control", In 2023 IEEE International Conference on Image Processing, 2023.

[24] M. J. Murdoch, M. E. Miller and P. J. Kane, “Perfecting the color reproduction of RGBW OLED”, Proceedings of the 30th International Conference on Imaging Science (ISIS), 2006.