Luminance Enhancement and Detail Preservation of Images and Videos Adapted to Ambient Illumination

Qing Song, *Member, IEEE*, and Pamela C. Cosman, *Fellow, IEEE*

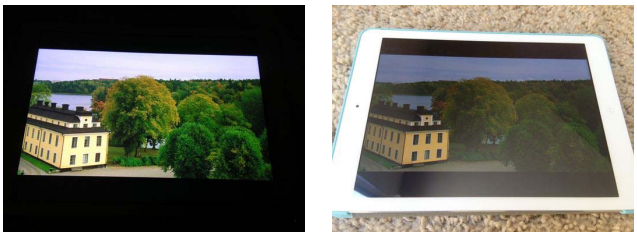
***Abstract*— When images and videos are displayed on a mobile device in bright ambient illumination, fewer details can be perceived than in the dark. The detail loss in dark areas of the images/videos is usually more severe. The reflected ambient light and the reduced sensitivity of viewer's eyes are the major factors. We propose two tone mapping operators to enhance the contrast and details in images/videos. One is content independent and thus can be applied to any image/video for the given device and the given ambient illumination. The other tone mapping operator uses simple statistics of the content. Display contrast and human visual adaptation are considered to construct the tone mapping operators. Both operators can be solved efficiently. Subjective tests and objective measurement show the improved quality achieved by the proposed methods.**

***Index Terms*— Ambient adaptation, luminance enhancement, tone mapping, contrast sensitivity.**

I. Introduction

THE number of users watching videos on mobile devices is growing rapidly. Mobile video traffic accounted for 60% of total mobile data traffic in 2016, and the percentage is expected to increase to 78% in the next five years [1]. Thanks to the convenience of mobile devices, users can enjoy videos indoors or outdoors, sitting or working out on a running machine. The viewing conditions, including display size and viewing distance, display brightness and ambient illumination, user movement, etc., affect the quality of experience of videos. Among them, ambient illumination varies greatly. A typical living room is 50 lx; a bright office can be 500 lx; outdoor under shade can be 5000 lx; an overcast day can be 10,000 lx; and under direct sunlight, it can be 100,000 lx. In a room with windows, the ambient illumination can vary from 0 to 1000 lx at different times of day.

When our eyes are adapted to bright ambient light, the amount of light that enters our eyes is affected, and thus the visual sensitivity is affected [2], [3]. In addition, the reflection of ambient light reduces the contrast of a display (contrast is defined in Sec. II). Therefore, fewer details can be perceived in bright surroundings than in dark. Moreover, the detail loss is more severe in dark areas in the video. Fig. 1 shows an image displayed in dark and in bright ambient. When displayed in the dark, the image looks bright, and shows good details. When displayed in bright ambient light, it loses details and contrast, and appears dull and washed out. Though increasing the display brightness can compensate for some of the detail loss, it will drain the battery of the device. Note that the maximum brightness of most mobile devices today is about 200 - 600 cd/m^2 [4]. In very bright ambient light (e.g., 10,000 lx), the quality of experience of viewing the display at even the maximum brightness is still not comparable to viewing in the dark.



(a) (b)

Fig. 1. An image displayed in different ambient illuminations. (a) In the dark. (b) In bright ambient.

Since details are less noticeable, some researchers proposed to compress videos using lower bit rate according to the ambient illumination. In [5], perceptual quality of video is predicted using a model of ambient illumination, bit rate, and structural and motion complexity of the video. The bit rate that achieves the same perceptual quality in the given bright ambient as the quality in low ambient illumination can be inversely derived. The extension of this work to 3D video is in [6] and [7]. In [8], the influence of display size and viewing distance, ambient illumination and body movement is modeled from the results of subjective tests. The perceptual distortion is predicted, and the compression of videos is adjusted accordingly. None of these works studied the greater effects of ambient light on the dark areas of videos than on the bright areas. Note that these approaches assume no tone mapping or enhancement of videos at the decoder side. If the contrast or the luminance of videos is enhanced at the decoder after decompression, the visibility of details can be improved, and the distortion caused by compression can become more noticeable.

Manuscript received May 1, 2017; revised February 2, 2018, March 24, 2018, and May 19, 2018; accepted May 30, 2018. Date of publication June 12, 2018; date of current version July 2, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Denis Kouame. *(Corresponding author: Qing Song.)*

Q. Song was with the University of California at San Diego, La Jolla, CA 92093 USA. She is now with Dolby Laboratories, Inc., Sunnyvale, CA 94085 USA (e-mail: qingsong@ieee.org).

P. C. Cosman is with the University of California at San Diego, La Jolla, CA 92093 USA (e-mail: pcosman@eng.ucsd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2846686

Some works have studied image/video enhancement to compensate for the effects of ambient light. Mantiuk *et al.* [9] constructed a tone mapping operator so that the human visual response of the enhanced image under bright ambient illumination can be as close to the maximum response as possible. The algorithm involves Laplacian decomposition and quadratic programming, and is complicated. In [10], images are enhanced by adjusting the backlight (screen brightness) to achieve the same visual response as in low ambient light. This method results in increasing the screen luminance for white. If the screen brightness for white is fixed, the method will result in clipping the bright areas of images. In [11], the tone mapping curve is constructed by establishing a linear relation between the display luminance and visual response. Kim [12] modeled an ambient-affected contrast sensitivity function, and designed an adaptive weighting filter in the spatial frequency domain. In [13], images are enhanced by boosting the gradients and improving the brightness by linear mapping. However, contents in bright areas are clipped, and the reflection of ambient light is not considered. Su *et al.* [14], [15] proposed to enhance luminance using an exponential function and to increase the gradients of the image, by taking into account both the reflection and visual sensitivity.

In this work, we propose two tone mapping operators to enhance the luminance and preserve the detail visibility of images and videos. One is content independent; the other uses some content statistics. Both need very light computation. They are built under the condition that the relationship between the amplitude level (codeword, or pixel value) and the display luminance is fixed, and the screen brightness for white is not allowed to increase. The device can detect the ambient illumination using its built-in ambient light sensor. If the computing resource of the device is very limited, the content independent tone mapping can be constructed for the given ambient illumination, and can be applied to any image/video. If a bit more computing resource is available, the content dependent tone mapping can be derived for each image/frame or group of frames, depending on how content statistics are used. No other image processing (e.g., gradient enhancement) is used.

The rest of the paper is organized as follows: The display and contrast models are explained in Sec. II, and the proposed tone mapping operators are described in Sec. III. Sec. IV shows the performance of the tone mapping and the comparisons with other tone mapping methods. Sec. V concludes the paper.

II. Display and Contrast Models

A. Display Luminance

According to [16], the electro-optical transfer function (EOTF) of a given 8-bit display is:

$$L_d(Y, L_W) = a(L_W)\big(\max[\frac{Y}{255} + b(L_W), 0]\big)^{\gamma},$$
$$\text{where} \quad a(L_W) = (L_W^{\frac{1}{\gamma}} - L_B(L_W)^{\frac{1}{\gamma}})^{\gamma},$$
$$b(L_W) = \frac{L_B(L_W)^{\frac{1}{\gamma}}}{L_W^{\frac{1}{\gamma}} - L_B(L_W)^{\frac{1}{\gamma}}}, \tag{1}$$

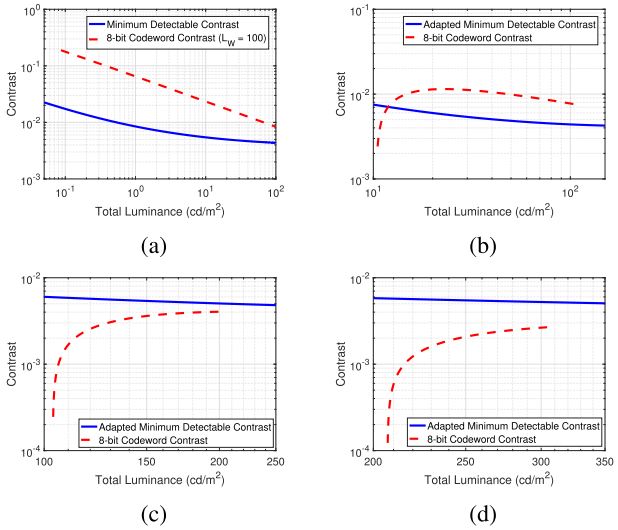


Fig. 2. Display codeword contrast and minimum detectable contrast of the ideal situation (no reflected light, full adaptation to each luminance) and adaptive minimum detectable contrast under ambient illumination 500 lx, 5000 lx and 10,000 lx for $L_W = 100$ cd/m^2. (a) $L_{refl} = 0$, full adaptation. (b) $E_{amb} = 500$ lx. (c) $E_{amb} = 5000$ lx. (d) $E_{amb} = 10000$ lx.

where L_d is the display luminance in cd/m^2, Y is the luma value (0-255) of a pixel, γ is a display gamma, L_W is the selected screen brightness for white in cd/m^2, and $L_B(L_W)$ is the screen brightness for black which is determined by L_W for a given device. L_W and L_B are non-negative, so (1) can be reduced to $L_d\ (Y, L_W) = a(L_W)\big(\frac{Y}{255} + b(L_W)\big)^{\gamma}$.

According to [9], the reflected light of the ambient illumination can be modeled as:

$$L_{refl}(E_{amb}) = \frac{k}{\pi} E_{amb}, \tag{2}$$

where k is the reflectivity of the display, and E_{amb} is the ambient illumination in lx. Lambertian reflection is assumed here. It is a conservative assumption, but is widely used in the literature [9], [10], [14], [15]. Specular reflection would reduce the dynamic range of the display even more, and have more severe impact on the viewing experience than Lambertian reflection.

The total luminance from a display is:

$$L_{total}(Y, L_W, E_{amb}) = L_d(Y, L_W) + L_{refl}(E_{amb})$$
$$= a(L_W)\big(\frac{Y}{255} + b(L_W)\big)^{\gamma} + \frac{k}{\pi} E_{amb}. \tag{3}$$

The contrast [17], [18] between each two consecutive codewords (namely, codeword contrast) is calculated as:

$$C_d(Y, L_W, E_{amb}) = 2\frac{L_{total}(Y+1, L_W, E_{amb}) - L_{total}(Y, L_W, E_{amb})}{L_{total}(Y+1, L_W, E_{amb}) + L_{total}(Y, L_W, E_{amb})}, \tag{4}$$

for $Y = 0, 1, \cdots, 254$. Fig. 2(a) shows the 8-bit display codeword contrast (dash red curve) when there is no reflected light from ambient illumination ($L_{refl} = 0$). The maximum screen brightness (L_W) is set to 100 cd/m^2. The display gamma is 2.23, and the reflectivity is 6.5%, which are the values for an iPad Air from the Display Mate website [4].

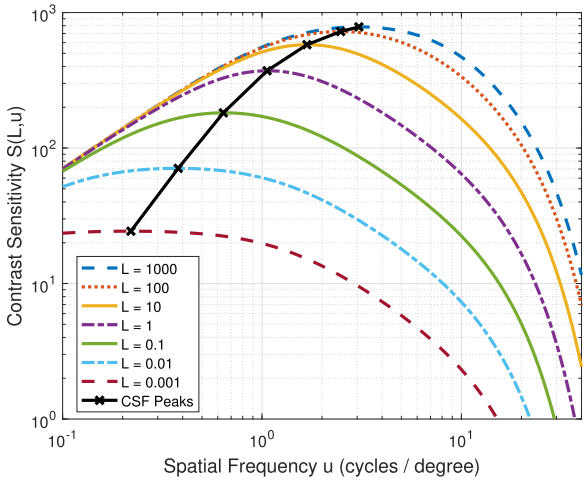


Fig. 3. Tracking the peaks of contrast sensitivity [20].

B. Minimum Detectable Contrast

According to the Weber-Fechner law, the minimum contrast that humans can detect is constant regardless of luminance. However, this law holds only for a mid-range of luminance values, and the measured minimum detectable contrast increases below and above certain threshold luminance values. According to [18], the luminance-dependent minimum detectable contrast is derived from Barten's contrast sensitivity function (CSF) [19]. Contrast sensitivity is defined as the inverse of the modulation threshold of a sinusoidal luminance pattern. The CSF at luminance L and frequency u is modeled in [19] as:

$$S(L,u)=\frac{e^{-2\pi^2\sigma^2u^2/\kappa}}{\sqrt{\frac{2}{T}\left(\frac{1}{X_o^2}+\frac{1}{X_{max}^2}+\frac{u^2}{N_{max}^2}\right)\left(\frac{1}{\eta pE}+\frac{\Phi_0}{1-e^{-(u/u_0)^2}}\right)}}, \quad (5)$$

where

$$\sigma = \sqrt{\sigma_0^2+(C_{ab}\,d)^2} \ \text{ arc min},$$
$$d = 5-3\tanh\left(0.4\log\ (L\,X_0^2/40^2)\right) \ \text{ mm},$$
$$E = \frac{\pi d^2}{4}L\left(1-(d/9.7)^2+(d/12.4)^4\right) \ \text{ Td},$$

and where $\kappa = 3$, $\sigma_0 = 0.5$ arc min, $u_0 = 7$ cycles/deg, $C_{ab} = 0.08$ arc min/mm, $X_{max} = 12°$, $T = 0.1$ sec, $N_{max} = 15$ cycles, $\eta = 0.03$, $\Phi_0 = 3\times10^{-8}$ sec deg^2, $p = 1.2\times10^6$ photons $\cdot$ sec$^{-1}\cdot$ deg$^{-2}\cdot$ Td^{-1}. X_o is usually set to 40°. To find the minimum detectable contrast at luminance L, the highest sensitivity is found over frequency [20]:

$$S_{max}(L)=\max_u\ S(L,u). \quad (6)$$

Fig. 3 shows the tracking of peaks of contrast sensitivity when adjusting luminance levels [20]. The minimum detectable contrast $C_t(L)$ for every luminance level is calculated in [20] as:

$$C_t(L)=\frac{1}{S_{max}(L)}\times\frac{2}{1.27}, \quad (7)$$

where the factor 2 is used for the conversion from modulation to contrast, and the factor $1/1.27$ is used for the conversion from sinusoidal to rectangular waves [18]. Fig. 2(a) shows $C_t(L)$ (solid blue curve) which is called the "Barten ramp" in [18] and [20]. Note that the CSF in (5) is for the scenario where human eyes are fully adapted to the luminance L.

C. Ambient-Affected Perceptual and Display Contrast

When the eyes are adapted to some other luminance L_s, the CSF model is modified in [19] as:

$$\tilde{S}(L,u,L_s)=S(L,u)\cdot e^{-\frac{\ln^2\left(\frac{L_s}{L}(1+\frac{144}{X_o^2})^{0.25}\right)-\ln^2\left((1+\frac{144}{X_o^2})^{0.25}\right)}{2\ln^2(32)}}. \quad (8)$$

Following the procedure of constructing the minimum detectable contrast for full adaptation, we construct the minimum detectable contrast when the eyes are adapted to L_s. The peaks of contrast sensitivity are found using:

$$\tilde{S}_{max}(L,L_s)=\max_u\ \tilde{S}(L,u,L_s). \quad (9)$$

The adapted minimum detectable contrast $\tilde{C}_t(L,L_s)$ is computed as:

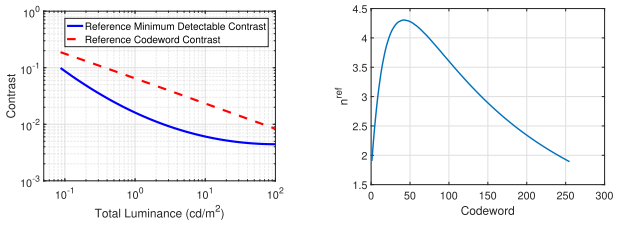
$$\tilde{C}_t(L,L_s)=\frac{1}{\tilde{S}_{max}(L,L_s)}\times\frac{2}{1.27}. \quad (10)$$

Figs. 2(b) - 2(d) show the adapted minimum detectable contrast when human eyes are adapted to the ambient illumination 500 lx (bright office), 5000 lx (outdoor in shade) and 10,000 lx (overcast day) where $L_s=\frac{E_{amb}}{\pi}$. The codeword contrast of 8-bit displays under the ambient illumination is also plotted using (4). As the ambient illumination gets brighter, the adapted minimum detectable contrast increases, and the codeword contrast decreases, thus the codeword contrast gradually drops below the adapted minimum detectable contrast. Note that the codeword contrast in Fig. 2 is plotted on logarithmic scales over the range of the total luminance, $[L_B(L_W)+\frac{k}{\pi}E_{amb}, L_W+\frac{k}{\pi}E_{amb}]$. Under 5000 lx and 10,000 lx, all the codeword contrasts are below the adapted minimum detectable contrast. A codeword contrast lower than the adapted minimum detectable contrast indicates the difference between a codeword and the next codeword cannot be perceived by human eyes. That results in the reduction or loss of perception of details in bright ambient light. The contrast of dark codewords drops more significantly, yielding more perception loss.

III. Proposed Luminance Enhancement

We want to enhance images/videos under bright ambient light so that more details become visible. The EOTF of the display is fixed, and the screen luminance for white (L_W) is pre-determined. Our goal is to find a tone mapping function to improve the contrast.

From the last section, it is known that the contrast of codewords decreases, especially for the dark codewords, when the ambient light gets brighter. Note that the contrast ratio of a display [17] is $\frac{L_{total}(255,L_W,E_{amb})}{L_{total}(0,L_W,E_{amb})}$, which is determined by L_W and E_{amb}. For a given L_W under a given E_{amb}, it is not possible to enhance the contrast of every codeword. The contrast of



(a) (b)

Fig. 4. Reference contrast and reference relative codeword contrast. (a) Reference contrast. (b) n^{ref}

dark codewords is reduced more than bright codewords under bright ambient illumination, so we will enhance the contrast of dark codewords. That means the contrast of some other codewords will be sacrificed.

The tone mapping function will be built so that the tone mapped codewords would have similar contrast distribution to that under the reference viewing condition. The following is defined as the reference viewing condition: the screen brightness for white (L_W^{ref}) is 100 cd/m^2 which is typical for 8-bit displays; the ambient is dark; and eyes are adapted to $L_s^{ref} = (L_W^{ref} + L_B(L_W^{ref}))/2 = (100 + L_B(100))/2$.

Using these settings, we compute the adapted minimum detectable contrast using (10) and the codeword contrast using (4), and plot them in Fig. 4(a). We define *relative codeword contrast* as the codeword contrast in the unit of the adapted minimum detectable contrast, which represents the number of just-noticeable-differences (JNDs) that the difference between each two consecutive codewords spans. We define the *reference relative codeword contrast* for each codeword as the relative codeword contrast under the reference viewing condition:

$$n^{ref}(Y, L_W^{ref}) = \frac{C_d(Y, L_W^{ref}, 0)}{\tilde{C}_t\big(L_{total}(Y, L_W^{ref}, 0), L_s^{ref}\big)}, \tag{11}$$

where $Y = 0, 1, \cdots, 254$. The reference relative codeword contrast is plotted in Fig. 4(b).

A. Content Independent Luminance Enhancement

A content independent tone mapping operator is proposed in this section. It does not use any content-related data, but only uses display characteristics (e.g., luminance for white L_W, gamma γ, and reflectivity k) and the ambient illumination. For a given display and a given ambient illumination, this tone mapping operator can be globally applied to any video or image. Very light computation is needed.

Using the reference relative codeword contrast n^{ref}, we propose to allocate codewords so that the relative codeword contrast after tone mapping under the given ambient light would be equal to $\alpha\, n^{ref}$, where α is positive. That is, the relative codeword contrast should satisfy:

$$\frac{C_d(T^G(Y,\alpha), L_W, E_{amb})}{\tilde{C}_t\Big(L_{total}\big(T^G(Y,\alpha), L_W, E_{amb}\big), \frac{E_{amb}}{\pi}\Big)} = \alpha\, n^{ref}(Y, L_W^{ref}), \tag{12}$$

where $T^G(Y,\alpha)$ is the output of tone mapping, i.e., $T^G(\cdot)$ is the (global) content independent tone mapping operator. Combining (4) and (12), we obtain:

$$\frac{2\frac{L_{total}\left(T^G(Y+1,\alpha),L_W,E_{amb}\right)-L_{total}\left(T^G(Y,\alpha),L_W,E_{amb}\right)}{L_{total}\left(T^G(Y+1,\alpha),L_W,E_{amb}\right)+L_{total}\left(T^G(Y,\alpha),L_W,E_{amb}\right)}}{\tilde{C}_t\Big(L_{total}\big(T^G(Y,\alpha), L_W, E_{amb}\big), \frac{E_{amb}}{\pi}\Big)} = \alpha\, n^{ref}(Y, L_W^{ref}), \tag{13}$$

The total luminance of $T^G(Y+1,\alpha)$ is obtained as a function of the total luminance of $T^G(Y,\alpha)$:

$$L_{total}(T^G(Y+1,\alpha), L_W, E_{amb}) = L_{total}\big(T^G(Y,\alpha), L_W, E_{amb}\big)\frac{2 + P(Y, L_W, E_{amb}, \alpha)}{2 - P(Y, L_W, E_{amb}, \alpha)},$$

where

$$P(Y, L_W, E_{amb}, \alpha) = \alpha\, n^{ref}(Y, L_W^{ref})\tilde{C}_t\Big(L_{total}\big(T^G(Y,\alpha), L_W, E_{amb}\big), \frac{E_{amb}}{\pi}\Big). \tag{14}$$

Therefore, the total luminance of each codeword can be derived recursively from that of the previous codeword. The inverse of the display luminance model (3) is then applied to compute the tone mapping operator:

$$T^G(Y,\alpha) = \Big(\big(\frac{L_{total}(T^G(Y,\alpha), L_W, E_{amb}) - \frac{k}{\pi}E_{amb}}{a(L_W)}\big)^{\frac{1}{\gamma}} - b(L_W)\Big)\cdot 255. \tag{15}$$

In most images and videos, a large portion of pixels are in the mid-tone. Therefore, very dark and very bright codewords can be clipped to further improve the contrast of mid-tones. That is, we set the total luminance of the codewords below z ($0 \le z < 255$) to the total luminance of black ($L_{total}(0, L_W, E_{amb})$). The total luminance of codewords in $[z, 255-z]$ is derived recursively from the previous codeword. The total luminance of the codewords above $255-z$ is set to $L_{total}(T^G(255-z,\alpha), L_W, E_{amb})$. In other words, a number of z codewords at both ends are clipped. A larger value of z would allow higher contrast of codewords in $[z, 255-z]$. $z = 20$ is recommended for 8-bit images/videos which means 15% codewords will be suppressed. In summary, the total luminance of codewords are computed as in (16), as shown at the bottom of the next page.

Now the problem is to find α. We want to enhance the relative codeword contrast as much as possible, i.e., we want α to be as large as possible. Note that if $L_{total}(T^G(255-z,\alpha), L_W, E_{amb})$ exceeds the selected luminance for white (L_W), more codewords would be clipped. In order to avoid that, we formulate the problem as:

$$\begin{aligned} &\max\ \alpha \\ &\text{s.t. } L_{total}(T^G(255-z,\alpha), L_W, E_{amb}) \le L_W \end{aligned} \tag{17}$$

The problem can be solved easily by bisection search. If the contrast ratio of the actual viewing condition is lower than the

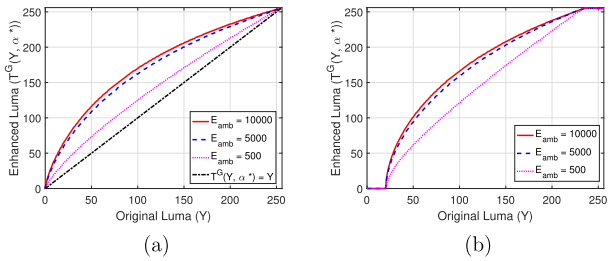


Fig. 5. Tone mapping curves for $L_W = 200$: $T^G(Y, \alpha^*)$ vs. Y. (a) $z = 0$. (b) $z = 20$.

contrast ratio of the reference condition, i.e.,

$$\frac{L_{total}(255, L_W, E_{amb})}{L_{total}(0, L_W, E_{amb})} < \frac{L_{total}(255, L_W^{ref}, 0)}{L_{total}(0, L_W^{ref}, 0)}, \quad (18)$$

then the optimum α would be less than 1. Fig. 5 shows the tone mapping curves for $L_W = 200$ and $E_{amb} = 500$, 5000 and 10,000, when z is 0 and 20. $T^G(Y, \alpha) = Y$ (linear mapping) is also plotted as a reference. Fig. 6 shows the codeword contrast and the adapted minimum detectable contrast before and after tone mapping. The relative codeword contrast after tone mapping is proportional to the reference relative codeword contrast. Compared to the codeword contrast before tone mapping, the contrast of dark codewords after tone mapping is enhanced, whereas the contrast of bright codewords is suppressed. In other words, the contrasts of codewords are re-allocated by the tone mapping. The contrast of dark codewords is more enhanced when the ambient illumination is higher.

Bright ambient light reduces the saturation of the video as well as the luminance contrast. We enhance the chrominance of the video by applying the simple method from [9]:

$$R(V) = \left(\frac{V}{Y}\right)^s T^G(Y, \alpha^*), \quad (19)$$

where V is the chroma value, $R(V)$ is the enhanced chroma value, and s is a constant.

Fig. 7 shows one frame from sequence *Park Scene* before and after tone mapping using the curve in Fig. 5(b) and (19) where $s = 0.8$. The images depict simulated appearance on the display, not the input to the display. The enhanced images look brighter and more saturated than the original image under the given ambient illumination, and more details in dark areas are revealed. As depicted in Fig. 6, the codeword contrast in bright surroundings is not comparable to that in dark surroundings. Therefore, the image viewed under brighter ambient light inevitably shows lower contrast even after enhancement. It is the ambient light that reduces the dynamic range of the image, and while the tone mapping can mitigate the degradation from ambient light, it cannot completely compensate for it.

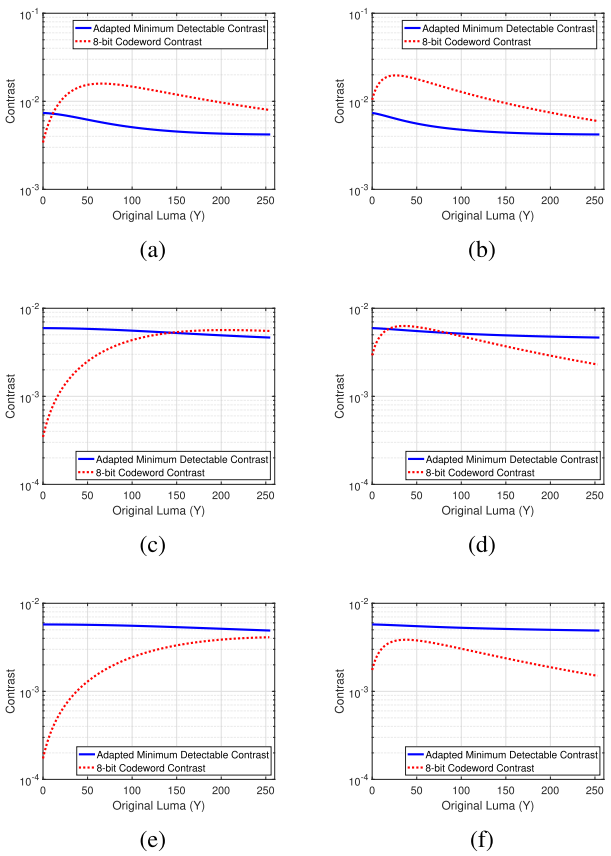


Fig. 6. Contrast before and after the proposed tone mapping for $L_W = 200$ and $z = 0$. (a) Contrast of original image: $E_{amb} = 500$. (b) Contrast of enhanced image: $E_{amb} = 500$. (c) Contrast of original image: $E_{amb} = 5000$. (d) Contrast of enhanced image: $E_{amb} = 5000$. (e) Contrast of original image: $E_{amb} = 10,000$. (f) Contrast of enhanced image: $E_{amb} = 10,000$.

B. Content Dependent Luminance Enhancement

In this section, we describe a content dependent tone mapping operator. In addition to the display characteristics and the ambient illumination, some statistics of the content are collected to construct the tone mapping operator, so that the image/video can be enhanced better. Unlike the method of Mantiuk *et al.* [9] where the Laplacian pyramid is employed and the contrast probabilities are computed for every luminance range and every frequency, we simply collect the histograms of codewords of the image/video.

For a frame in a video, the histograms of codewords are collected as:

$$h_{f,m} = |\Phi_{f,m}|,$$
$$\text{where} \quad \Phi_{f,m} = \{j | \frac{256m}{M} \le I_{f,j} < \frac{256(m+1)}{M}\}, \quad (20)$$

and where $I_{f,j}$ is the j-th pixel in frame f, and $m = 0, 1, \cdots, M-1$. We set M to 32, i.e., there are 32 bins in

$$L_{total}(T^G(Y,\alpha), L_W, E_{amb}) = \begin{cases} L_{total}(0, L_W, E_{amb}) & \text{if } Y < z \\ L_{total}\big(T^G(Y-1,\alpha), L_W, E_{amb}\big)\dfrac{2 + P(Y-1, L_W, E_{amb}, \alpha)}{2 - P(Y-1, L_W, E_{amb}, \alpha)} & \text{else if } z \le Y \le 255 - z \\ L_{total}(T^G(255-z,\alpha), L_W, E_{amb}) & \text{otherwise,} \end{cases} \quad (16)$$

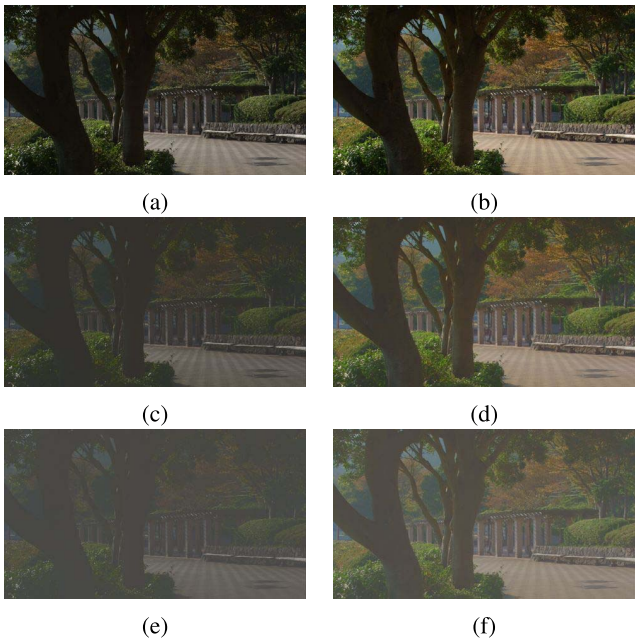


Fig. 7. Images before and after the proposed tone mapping for $L_W = 200$. Note that these are simulated appearance on the display, which however may not convey the actual brightness or color due to print limitation. (a) Original at 500 lx. (b) Enhanced at 500 lx. (c) Original at 5000 lx. (d) Enhanced at 5000 lx. (e) Original at 10,000 lx. (f) Enhanced at 10,000 lx.

the histograms. In order to produce smooth tone mapping curves, at least 16 bins are recommended for the construction of histograms for 8-bit images/videos.

We average the histograms of frames from f_x to f_y, and compute the weighting factor for codeword Y as:

$$w(Y) = \left(\frac{1}{f_y - f_x + 1}\sum_{i=f_x}^{f_y}\left(\frac{h_{i,b}}{\sum_{m=0}^{M-1} h_{i,m}}\right)\right)^{\beta},$$
$$\text{where} \quad b = \lfloor \frac{Y}{256} M \rfloor, \tag{21}$$

where β is a constant ($0 < \beta \leq 1$), so $w(Y)$ is in $[0, 1]$. The frames from f_x to f_y can correspond to a scene or a sliding window which includes the frame f. For example, f_x can be $f - 9$ and f_y can be f, and thus the histograms of 10 frames are averaged to give the weighting factors.

For single image enhancement, the histograms of the image are collected, i.e., $f_x = f_y = 1$, and the weighting factor is computed as (21).

The tone mapping operator is constructed using the weighted reference relative codeword contrast:

$$\frac{C_d(T^D(Y,\alpha), L_W, E_{amb})}{\tilde{C}_t\Big(L_{total}\big(T^D(Y,\alpha), L_W, E_{amb}\big), \frac{E_{amb}}{\pi}\Big)} = \alpha\, w(Y) \cdot n^{ref}(Y, L_W^{ref}), \tag{22}$$

where $T^D(\cdot)$ is the content dependent tone mapping operator. Compared to (12), the codewords are allocated so that the relative codeword contrast satisfies $\alpha\, w(Y) \cdot n^{ref}(Y, L_W^{ref})$ instead of $\alpha\, n^{ref}(Y, L_W^{ref})$. The contrast of the codewords corresponding to higher histogram counts is enhanced more, because these codewords take up larger areas in the video. For example, say the codewords under 50 take up 80% of a video. They are likely to be more perceptually important than the other codewords. The contrast of the codewords under 50 are more enhanced by multiplying the reference relative codeword contrast by larger weighting factors.

The contrast between two consecutive codewords after tone mapping is hence constructed in (23), as shown at the bottom of this page, by combining (4) and (22). We obtain the total luminance of $T^D(Y+1, \alpha)$ as in (24), as shown at the bottom of this page.

$T^D(0, \alpha)$ is set to 0, i.e., the total luminance of the first codeword, $L_{total}(T^D(0,\alpha), L_W, E_{amb})$, is set to $L_{total}(0, L_W, E_{amb}) = L_B(L_W) + \frac{k}{\pi}E_{amb}$. The total luminance of other codewords is derived recursively from that of the previous codeword. The problem is formulated similarly to (17):

$$\begin{aligned} &\max \ \alpha \\ &\text{s.t. } L_{total}(T^D(255,\alpha), L_W, E_{amb}) \leq L_W \end{aligned} \tag{25}$$

The tone mapping operator, $T^D(Y, \alpha^*)$, is then obtained by applying (15) where $T^G(Y, \alpha^*)$ is replaced by $T^D(Y, \alpha^*)$. The chrominance of the video is enhanced similarly to (19): $R(V) = \left(\frac{V}{Y}\right)^s T^D(Y, \alpha^*)$.

Figs. 8(a) - 8(d) show the weighting factors and the tone mapping curves for the original frame of Fig. 7 when $\beta = 0.4$ and $\beta = 1$. Figs. 8(e) and (f) show the corresponding simulated appearance of the enhanced images when E_{amb} is 5000 lx. The tone mapping curve of $\beta = 1$ is steeper than the curve of $\beta = 0.4$ for codewords between 25 and 150. More pixels are mapped to brighter codewords, as depicted in Fig. 8(j). Therefore, the enhanced image looks brighter. However, the tone mapping curve of $\beta = 1$ is almost flat for the

$$2\frac{L_{total}\big(T^D(Y+1,\alpha), L_W, E_{amb}\big) - L_{total}\big(T^D(Y,\alpha), L_W, E_{amb}\big)}{L_{total}\big(T^D(Y+1,\alpha), L_W, E_{amb}\big) + L_{total}\big(T^D(Y,\alpha), L_W, E_{amb}\big)} = \alpha\, w(Y) \cdot n^{ref}(Y, L_W^{ref}) \cdot \tilde{C}_t\Big(L_{total}\big(T^D(Y,\alpha), L_W, E_{amb}\big), \frac{E_{amb}}{\pi}\Big), \tag{23}$$

$$L_{total}(T^D(Y+1,\alpha), L_W, E_{amb}) = L_{total}\big(T^D(Y,\alpha), L_W, E_{amb}\big)\frac{2 + Q(Y, L_W, E_{amb}, \alpha)}{2 - Q(Y, L_W, E_{amb}, \alpha)},$$
$$\text{where} \quad Q(Y, L_W, E_{amb}, \alpha) = \alpha\, w(Y)\, n^{ref}(Y, L_W^{ref})\tilde{C}_t\Big(L_{total}\big(T^D(Y,\alpha), L_W, E_{amb}\big), \frac{E_{amb}}{\pi}\Big). \tag{24}$$

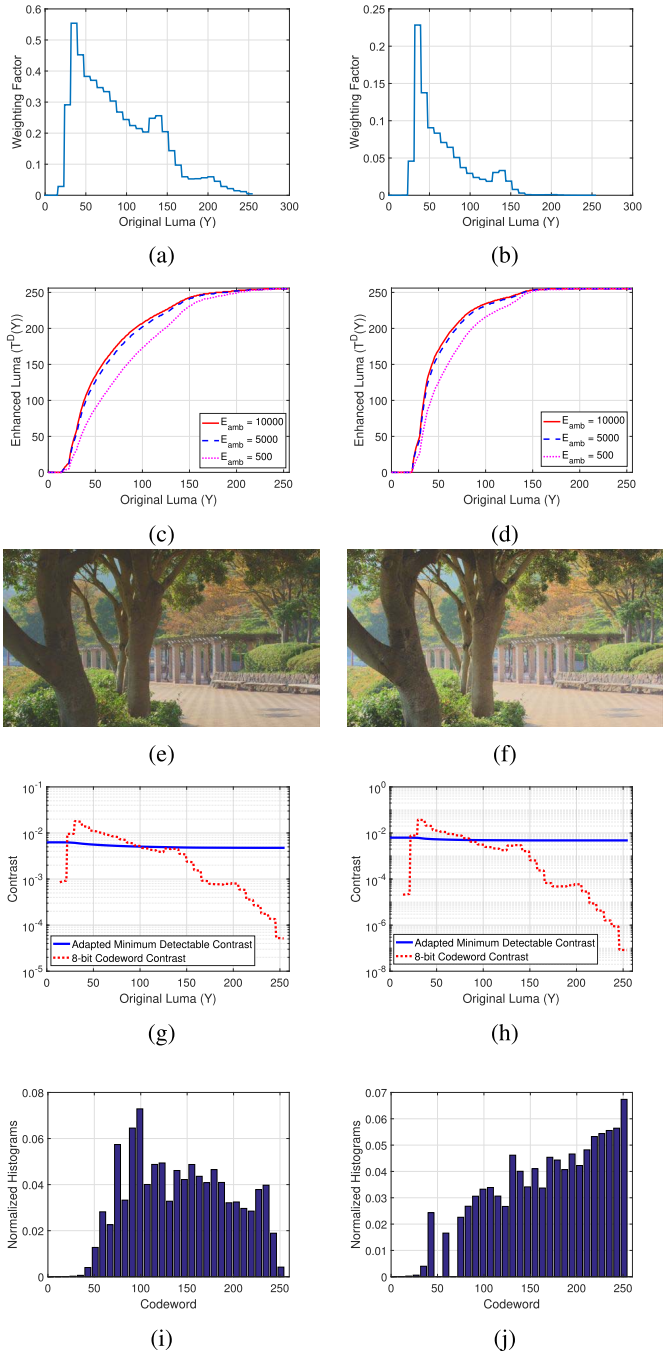


Fig. 8. Weighting factors, tone mapping curves, simulated appearance, contrast and normalized histograms of enhanced images for $E_{amb} = 5000$ when β is 0.4 and 1. (a) w(Y): $\beta = 0.4$. (b) w(Y): $\beta = 1$. (c) Tone mapping: $\beta = 0.4$. (d) Tone mapping: $\beta = 1$. (e) Enhanced image: $\beta = 0.4$. (f) Enhanced image: $\beta = 1$. (g) Contrast of enhanced image: $\beta = 0.4$. (h) Contrast of enhanced image: $\beta = 1$. (i) Normalized histograms of enhanced image: $\beta = 0.4$. (j) Normalized histograms of enhanced image: $\beta = 1$.

codewords above 150, thus yielding detail loss in bright areas in Fig. 8(f) (the texture on the ground is removed). The reason is that the weighting factors of those bright codewords are much smaller than the weighting factors of dark codewords, and thus the contrast of the bright codewords is ignored in the optimization. The heuristic value 0.4 is recommended for β.

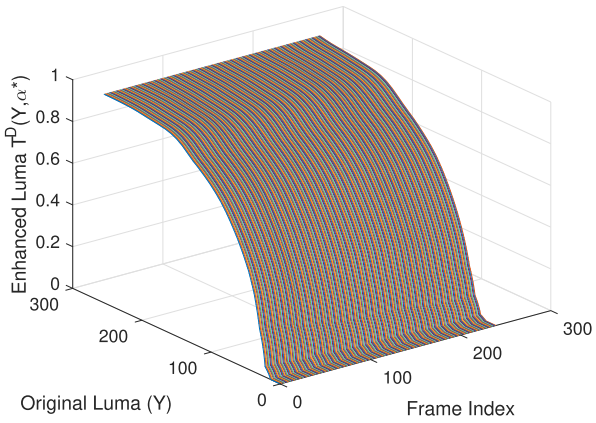


Fig. 9. Tone mapping curves of all the frames in the video: $E_{amb} = 5000$ and $\beta = 0.4$.

(a) (b) (c)

Fig. 10. Original images. (a) *Crowd Run*. (b) *Park Scene*. (c) *Kimono*.

When β is 0.4, the variance of the weighting factors is reduced, and thus the bright codewords have more impact on the tone mapping curve. As a result, the details in the bright regions are preserved better. Note that the codewords which correspond to short bins in the histograms should not be ignored completely, because they can be the foreground in the picture thus drawing the viewer's attention.

Compared to the content independent luminance enhancement, the contrast of codewords around 50 is boosted higher by the content dependent method, while the contrast of bright codewords is suppressed more. As a result, the enhanced images look brighter than Fig. 7(d). Codewords at both ends are clipped according to the histograms rather than selected heuristically.

We apply this method to the video sequence containing the original frame of Fig. 7. The tone mapping curves of all the frames in the sequence are plotted in Fig. 9, where $\beta = 0.4$, $f_y = f$ and

$$f_x = \begin{cases} 0 & f < 9 \\ f - 9 & \text{otherwise} \end{cases} \tag{26}$$

The tone mapping curves are all very similar to each other, and therefore, all the frames are enhanced similarly. The enhanced video is temporally stable.

Note for both proposed methods, L_W and $L_B(L_W)$ are needed to generate the tone mapping curves. The display device should have access to the values of L_W and $L_B(L_W)$. Since the proposed tone mapping is intended to be implemented in the display device, $L_B(L_W)$ should be available for each selected L_W.

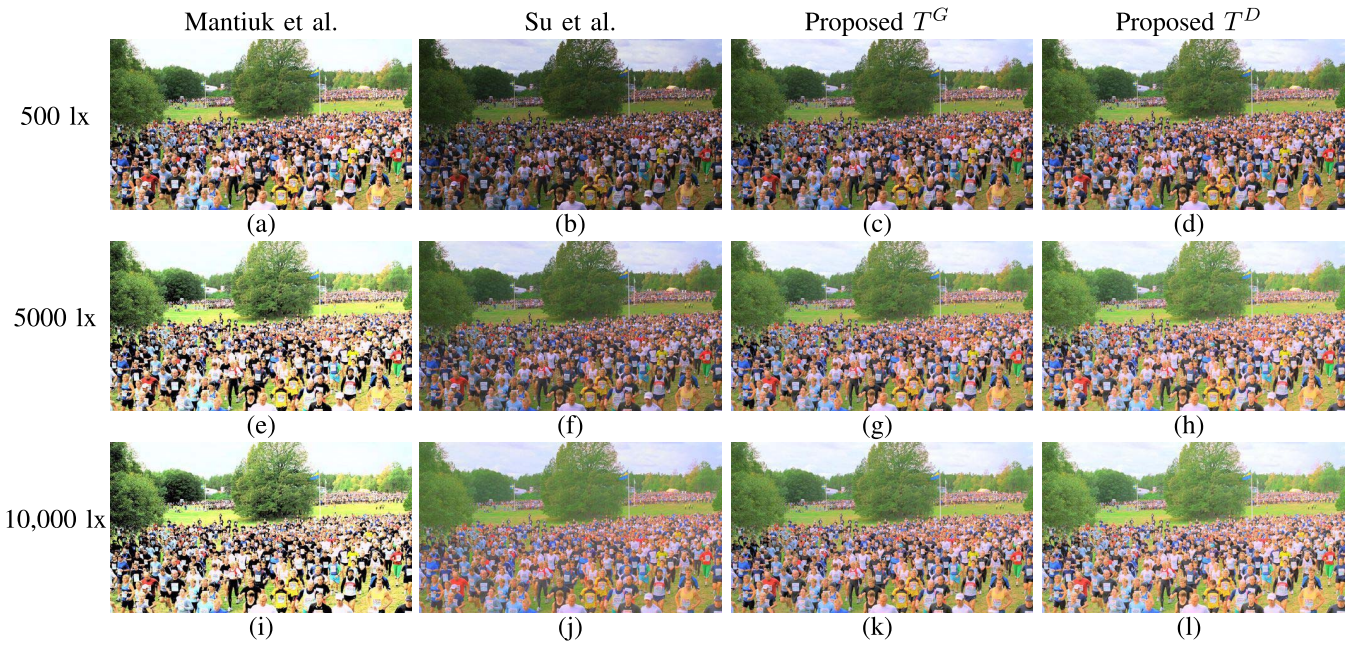


Fig. 11. *Crowd Run*: enhanced images using different algorithms for $L_W = 200$. Note that these are the inputs to the device, which do not show the look under the given ambient illumination.

IV. Performance Evaluation

A. Subjective Evaluation

We evaluate the performance of the tone mapping operators by a subjective test. Six randomly selected video clips (1920×1080) from [21] and [22] are enhanced using our proposed methods, the tone mapping method of Mantiuk *et al.* [9], and the adaptive luminance enhancement method of Su *et al.* [14], [15]. The content of each video does not change dramatically. We did two experiments. In the first experiment, subjects watched images extracted from videos in order to evaluate the luminance enhancement and detail preservation of the proposed methods. One image is extracted from each video. In the second experiment, subjects watched videos processed by the proposed content dependent method to check whether there is temporal inconsistency in the enhanced videos. The tone mapping generated by the proposed content independent method is global for all videos, so it cannot have any temporal issue.

The main focus of this work is on luminance enhancement, not on chrominance adjustment. To rule out the effects of different chrominance adjustment methods, chroma is enhanced using the same method in (19) from [9] for all the tone mapping methods, where $T^G(Y, \alpha^*)$ is replaced by each tone mapping operator. Some enhanced images are shown in Figs. 11-13. The corresponding original images are shown in Fig. 10. Note they are inputs to the device which do not show the appearance under the given ambient light.

The images and videos were shown on an iPad Air. The reflectivity k is 6.5%, the display gamma γ is 2.23, and the screen brightness for white L_W is adjustable from 6 to 449 cd/m^2 [4]. The size of the device is 9.7 inch. The viewing distance is 3 times the picture height. The experiment was conducted in three locations: in a bright office where the ambient illumination is 500 lx, in front of a building (under shade) where the ambient illumination is 5000 lx, and outdoor during an overcast day without shade or in a sunny day with slight shade where the ambient illumination is 10,000 lx. We used a lux meter to measure the ambient illumination to ensure it was what we assumed.

Four viewing conditions were tested: 1)

1) L_W is 200 cd/m^2 and E_{amb} is 500 lx,
2) L_W is 200 cd/m^2 and E_{amb} is 5000 lx,
3) L_W is 200 cd/m^2 and E_{amb} is 10,000 lx,
4) L_W is 449 cd/m^2 and E_{amb} is 10,000 lx.

In the experiment, because we could not access the built-in values of $L_B(L_W)$, we measure the values by a luminance meter. If neither the luminance meter nor the built-in values are available, $L_B(L_W)$ could be estimated by $\frac{L_W}{\text{contrast ratio}}$, where the contrast ratio is available on the display device official website and/or displaymate website [4]. Since the value of L_B is very small compared to L_W, a small error in the estimation can be neglected.

Twenty-eight subjects (age 18 - 53) participated in the subjective tests. Among them, 24 subjects are in their 20s. Ten subjects are female, and 18 are male. All of the subjects are non-experts and were unfamiliar with the content.

1) Image Comparison: Pair comparison [23], [24] is used to evaluate performance. Subjects compared images in pairs: one image processed by one of our proposed methods, and the other by one of the baseline schemes which include the method of Mantiuk et al., the method of Su et al., and no processing (the original image). The two proposed methods are also compared with each other. The images were labeled A and B randomly. The image pairs were shown sequentially on the same device. Subjects were requested to determine their preferred image in each pair. Five options were

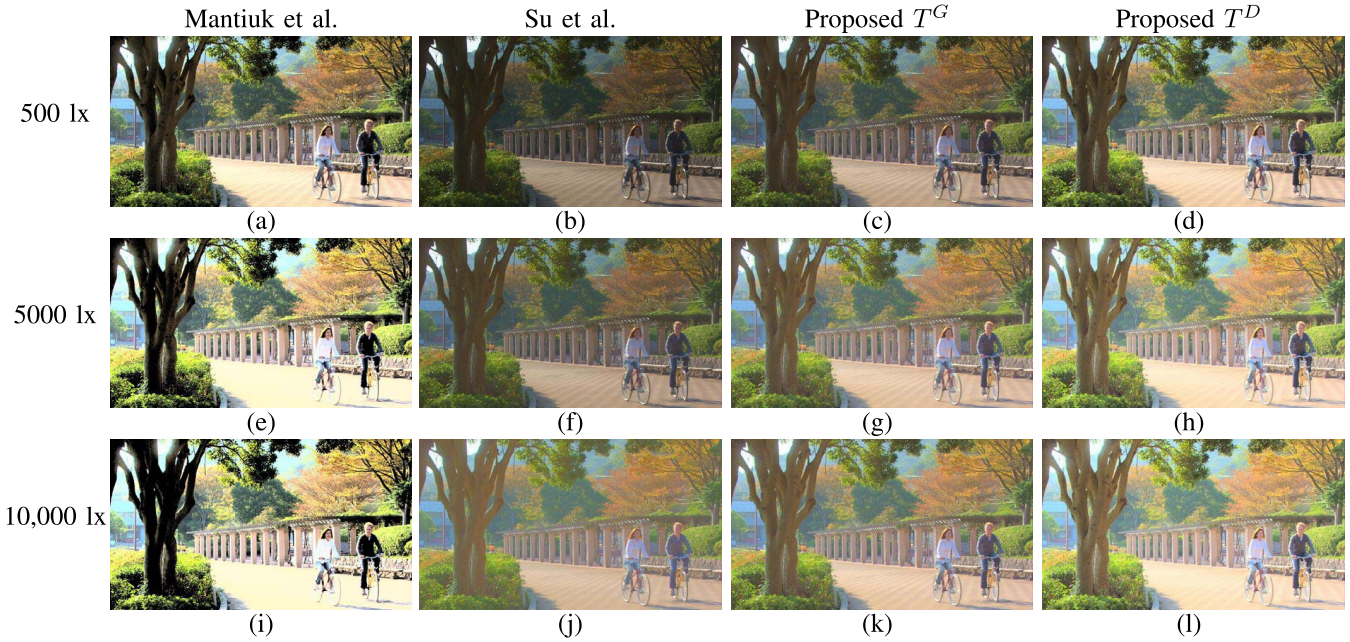


Fig. 12. *Park Scene*: enhanced images using different algorithms for $L_W = 200$. Note that these are the inputs to the device, which do not show the look under the given ambient illumination.

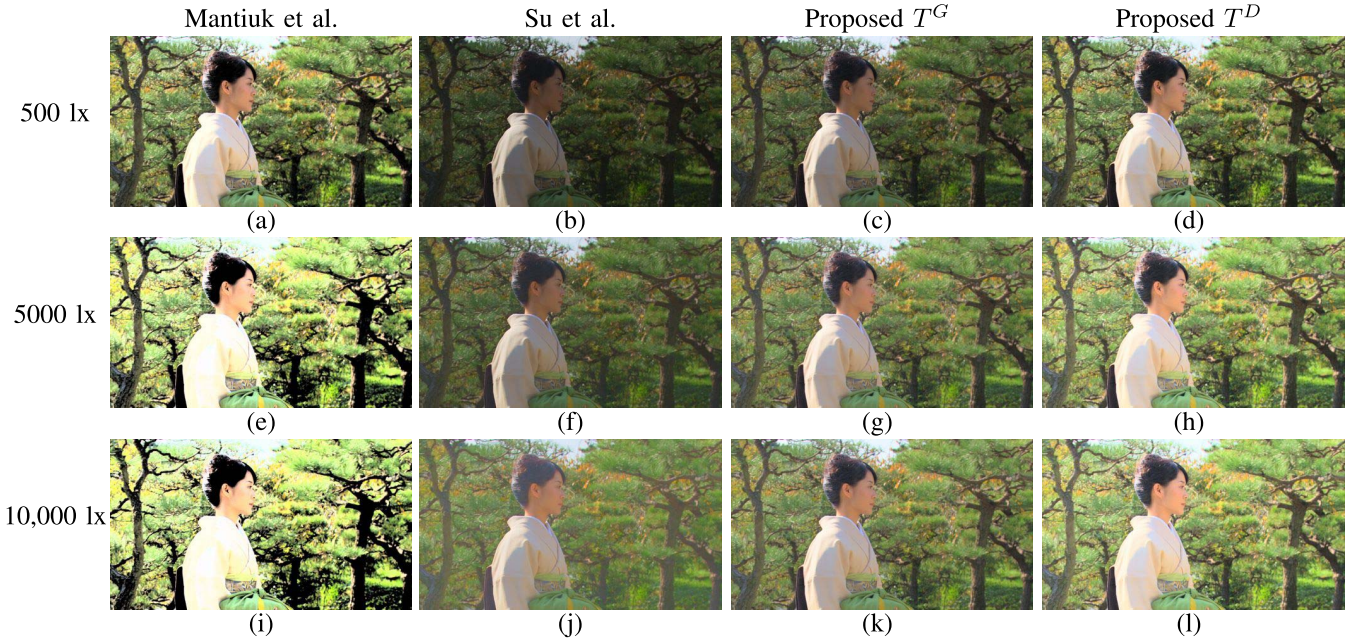


Fig. 13. *Kimono*: enhanced images using different algorithms for $L_W = 200$. Note that these are the inputs to the device, which do not show the look under the given ambient illumination.

given to subjects: "A is much better than B", "A is slightly better than B", "A is the same as B", "A is slightly worse than B", and "A is much worse than B". Subjects were also asked to select the reasons why they prefer one to the other one. The possible reasons were: being brighter, being darker, more details, higher contrast, or lower contrast.

We did not adopt the widely known forced choice (2-level rating) because it is known that when the items being compared are very similar, subjects become frustrated being forced to make a decision, and the frustration can impact the entire experiment. The 3-level rating ($A > B$, $A = B$, $A < B$) avoids the frustration problem as subjects find it easier to rate without guessing on invisible differences when they could not find any. The 5-level rating (a typical Likert scale [25]) that we used is similar to the 3-level rating, but we ask the subjects to provide a bit more information: how much difference is in each pair. We did this because the differences in some pairs of images are much more visible than the differences of other pairs. For example, the difference between the images before and after tone mapping is quite large under 10,000 lx.

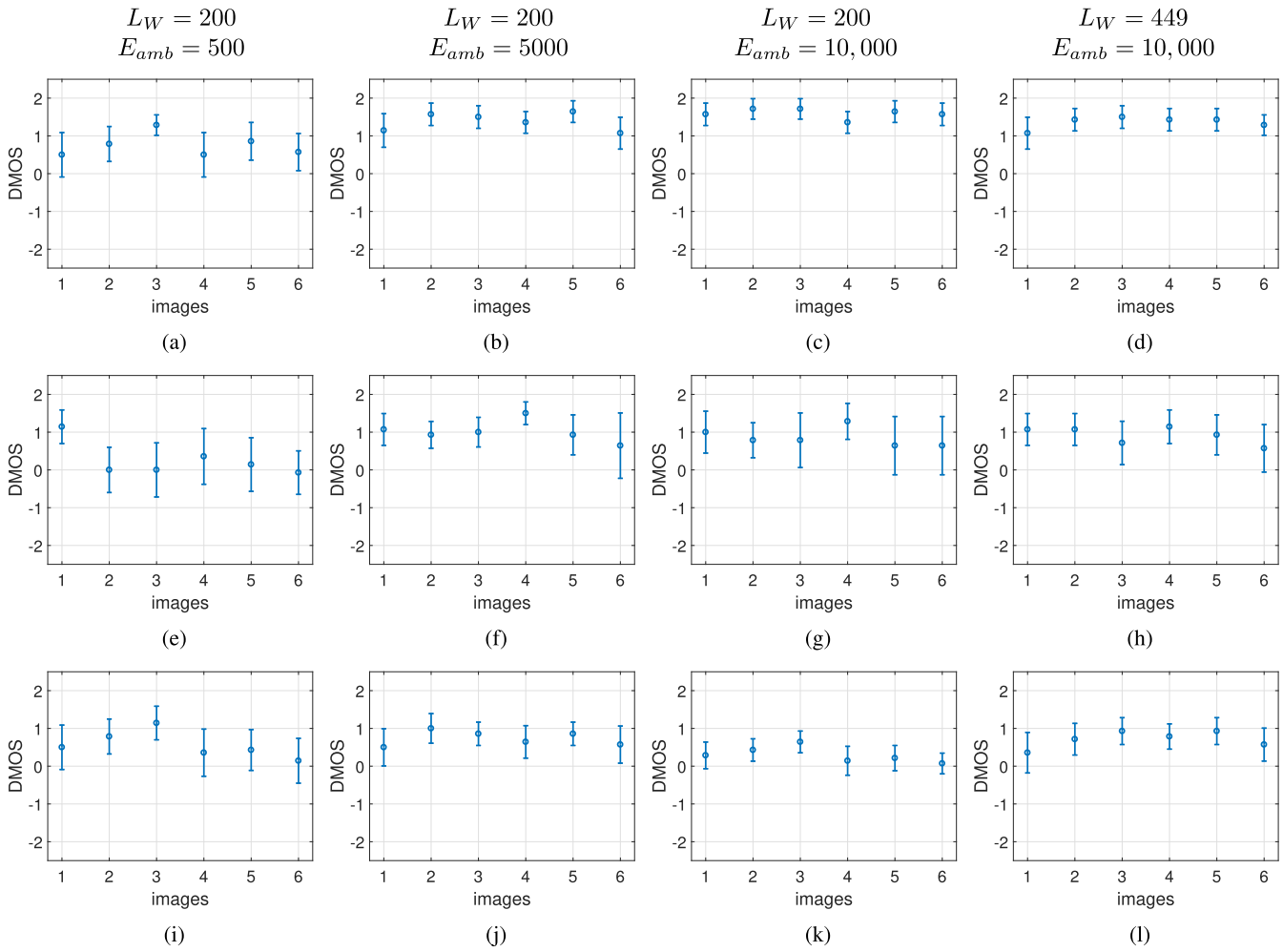


Fig. 14. 95% confidence intervals of DMOS of proposed content independent method vs. other schemes. (a) - (d): against no processing (original), (e) - (h): against the method of Mantiuk et al., (i) - (l): against the method of Su et al. Images 1 - 6 are *Crowd Run*, *Into Tree*, *Kimono*, *Old Town Cross*, *Park Scene* and *Rush Hour*.

However, the difference between the images processed by different tone mapping operators is smaller. If the 3-level rating is used, subjects might neglect the small differences between different tone mapping operators, and only look for and report large differences. The 5-level rating urges the subjects to pay more attention to all the differences they can perceive, so it is suitable when the experiment includes multiple competing processing approaches as well as unprocessed images.

a) Results of content independent enhancement: Fourteen subjects conducted the comparisons between the proposed content independent method and the baseline schemes. When our proposed method was rated much better (or worse) than the other scheme, the opinion score is $+2$ (or -2); when our proposed method was rated slightly better (or worse) than the other scheme, the opinion score is $+1$ (or -1); when no difference was found, the opinion score is 0. The difference mean opinion score (DMOS) is computed between the proposed method and the other schemes. Positive (or negative) numbers mean the proposed luminance enhancement works better (or worse) than the other scheme. We plot the DMOS and the 95% confidence intervals (CIs) in Fig. 14, where the first row is the DMOS against no processing, the second is against the method of Mantiuk et al., and the third is against the method of Su et al. Each column corresponds to one tested viewing condition. The average DMOS of all the images and CIs are plotted in Fig. 15, where "original" means no processing. CIs including zero indicate that we cannot reject the null hypothesis that the two schemes perform the same, or at least there is no consensus of preference.

When L_W is 200 cd/m^2, the enhanced images of the proposed content independent method have higher gain over the original images as the ambient light gets brighter. Under 500 lx (Fig. 14(a)), the CIs of 4 out of 6 images are above zero; while under 5000 lx and 10,000 lx (Figs. 14(b) and 14(c)), the CIs of all the 6 images are above zero. The average DMOS over images are 0.75, 1.38 and 1.60 for 500 lx, 5000 lx and 10,000 lx, respectively. Even when L_W is set to the maximum screen brightness 449 cd/m^2, the proposed scheme outperforms no processing by an average DMOS of 1.36, under 10,000 lx. The original images look dark and dull with low contrast under bright ambient light, and details are invisible. The proposed method improves the visibility, and enhances the brightness of images.

The method of Mantiuk et al. generally boosts the contrast of the mid-tone higher than our proposed method, which yields brighter images and sharper edges, but sometimes removes details in bright areas of images. For example, the clouds in Fig. 11(a, e, i), the details of the ground in Fig. 12(e, i), and the clothing shades in Fig. 13(e, i) are removed. More details are lost when the ambient light is brighter. Our proposed method (Figs. 11 - 13(c, g, k)) preserves those details very well.

The preference between the method of Mantiuk et al. and our proposed content independent method is controversial under 500 lx. The CIs are wide and include zero for 5 out of 6 images in Fig. 14(e). Subjects have various preferences in contrast and details. The proposed method shows clear advantage for image *Crowd Run*, as most subjects valued the details preserved by the proposed method.

The proposed method is favored over the method of Mantiuk et al. for most images under 5000 lx, due to more details and lower contrast. The advantage drops slightly when the ambient increases to 10,000 lx with L_W kept to 200 cd/m^2, because the favor of subjects shifts toward high contrast under such bright ambient where details in most images are hardly detectable. However, when L_W is turned up to 449 cd/m^2, our method which has higher visibility of details wins the comparison. When we pool the DMOS of all the test images (Fig. 15), the proposed method outperforms the method of Mantiuk et al. in all the viewing conditions, among which the DMOS of 5000 lx is the highest.

The results of the method of Su et al. are generally darker than the proposed content independent method, and show lower contrast. Under 500 lx where the eyes of subjects are relatively sensitive, the proposed method outperforms the method of Su et al. for 2 out of the 6 images (Fig. 14(i)). For example, the texture of the trees in Fig. 13(b) is undetectable and looks flat, whereas the result of our proposed method (Fig. 13(c)) shows the texture clearly. As a result, the DMOS between our proposed method and the method of Su et al. for this image is 1.14 under 500 lx, and the CI is well above zero (Fig. 14(i)).

Under 5000 lx, the proposed method outperforms the method of Su et al. for all the images (Fig. 14(j)). For example, Fig. 12(f) looks washed out under 5000 lx. Details in dark areas are still invisible in the bright surrounding as in the original image Fig. 10(b). Under 10,000 lx, the two methods perform similarly for 4 out of 6 images when L_W is 200 cd/m^2 (Fig. 14(k)), but the proposed method shows clear advantage when L_W is 449 cd/m^2 (Fig. 14(l)). When we pool all images in Fig. 15, the CIs of the average DMOS of all the viewing conditions are above zero, indicating the superiority of our method.

b) Results of content dependent enhancement: The proposed content dependent method and the other schemes were evaluated by another fourteen subjects. The DMOS and CIs of each image are plotted in Fig. 16. Each row corresponds to the DMOS against no processing, the method of Mantiuk et al., the method of Su et al., and the proposed content independent method. The images enhanced by the proposed content dependent method generally looks brighter and shows higher contrast than those enhanced by

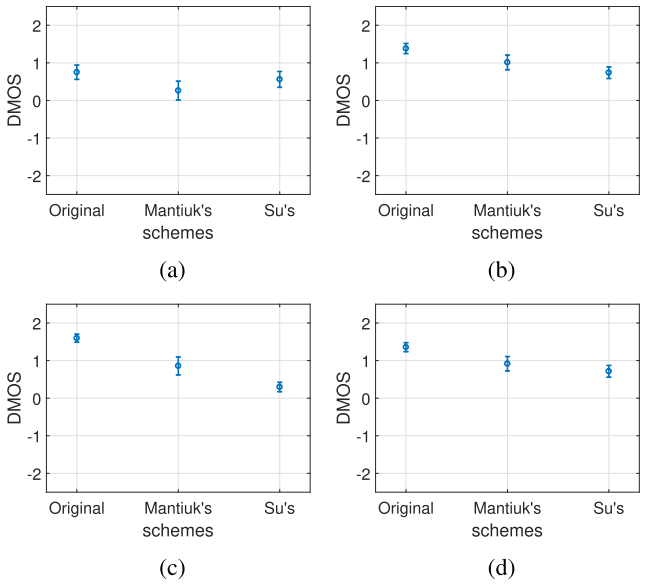


Fig. 15. 95% confidence intervals of average DMOS of proposed content independent method vs. other schemes for all images. (a) $L_W = 200$, $E_{amb} = 500$. (b) $L_W = 200$, $E_{amb} = 5000$. (c) $L_W = 200$, $E_{amb} = 10,000$. (d) $L_W = 449$, $E_{amb} = 10,000$.

the proposed content independent method. That is because the content dependent method puts more emphasis on the codewords which take up larger areas of the picture. The contrasts of those codewords are enhanced more than the other codewords. The difference in the image *Kimono* (Figs.13(c, g, k) and 13(d, h, l)) is the most obvious, and thus the DMOS of that image is the highest among all the test images under most viewing conditions. The CIs of the average DMOS for all the images against the proposed content independent scheme are all above zero (Fig. 17 where "Proposed global" means the proposed content independent method).

Like the proposed content independent method, the content dependent method greatly outperforms no processing, and the advantage grows with the increase of the ambient illumination. Under 500 lx (Fig. 16(a)), the proposed content dependent method is preferred for 4 out of 6 test images; under 5000 lx (Fig. 16(b)), it is favored for 5 out of 6 images; and under 10,000 lx (Figs. 16(c) - 16(d)), it beats no processing for all the images.

Under 500 lx, the proposed content dependent method performs similarly to the method of Mantiuk et al (Fig. 16(e)). The proposed method wins when the ambient light is brighter, because it shows a better trade-off between detail preservation and contrast enhancement. The CIs of 5 out of 6 images are above zero when the ambient illumination is 5000 lx and 10,000 lx (Figs. 16(f) - 16(h)).

The subjects preferred the proposed content dependent method to the method of Su et al. for most of the test images (Figs. 16(i) - 16(l)). The superiority is quite clear. The average DMOS of the four viewing conditions are 0.90, 0.93, 0.82 and 0.90.

In summary, both of the proposed methods outperform the baseline schemes. They keep details of images better than

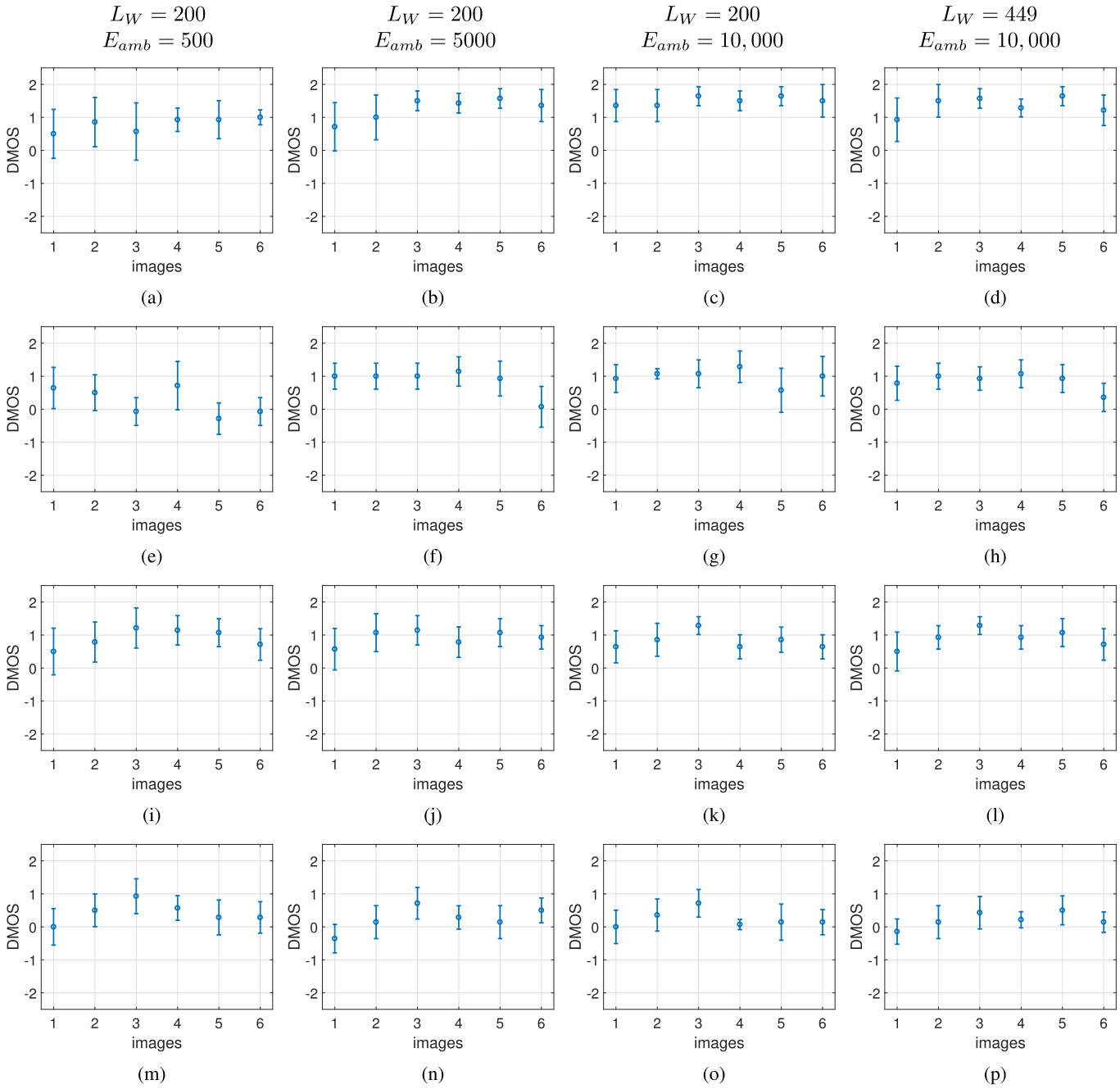


Fig. 16. 95% confidence intervals of DMOS of proposed content dependent method vs. other schemes. (a) - (d): against no processing (original), (e) - (h): against the method of Mantiuk et al., (i) - (l): against the method of Su et al., (m) - (p): against the proposed content independent method. Images 1 - 6 are *Crowd Run*, *Into Tree*, *Kimono*, *Old Town Cross*, *Park Scene* and *Rush Hour*.

the method of Mantiuk et al., and they improve the contrast and brightness more than the method of Su et al. Therefore, they win the comparisons when the ambient light is very bright. The advantages of both proposed methods over no processing are quite large, as the proposed methods enhance the brightness and contrast. The proposed content dependent method is slightly better than the content independent method, because the contrast of codewords with higher histogram counts is boosted higher.

Note that both of our proposed methods, content independent and dependent enhancement, are computationally much simpler than the method of Mantiuk et al. which is a content dependent approach. As stated in [9], half of the processing time of the method of Mantiuk et al. is spent on computing the contrast probabilities of each luminance range of each frequency of the Laplacian pyramid. The other half is spent on solving their optimization problem iteratively. Our content dependent method only collects histograms of codewords and does not do any frequency decomposition. The optimization time is also much shorter than the method of Mantiuk et al. The proposed content independent method is computationally even more efficient, as no video-related data is needed.

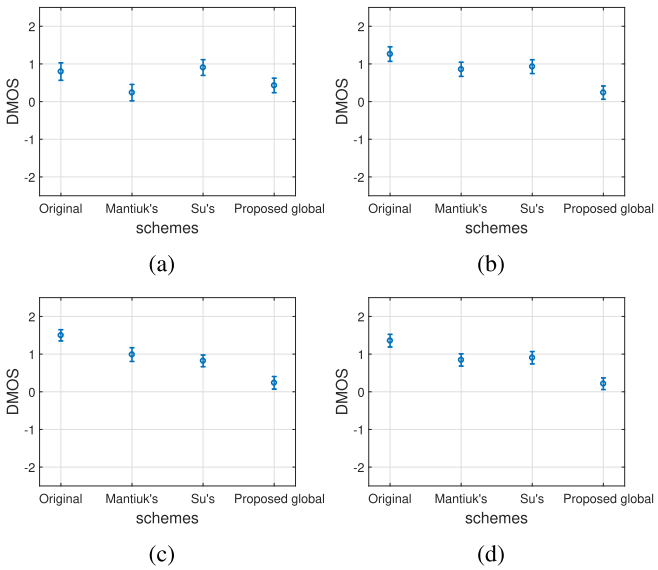


Fig. 17. 95% confidence intervals of average DMOS of proposed content dependent method vs. other schemes for all test images. "Proposed global" means the proposed content independent method. (a) $L_W = 200$, $E_{amb} = 500$. (b) $L_W = 200$, $E_{amb} = 5000$. (c) $L_W = 200$, $E_{amb} = 10,000$. (d) $L_W = 449$, $E_{amb} = 10,000$.

2) Video Comparison: To demonstrate the temporal stability of the proposed content dependent enhancement method, we had subjects compare the enhanced videos with the original videos in pairs. The videos were labeled A and B randomly. Subjects were asked to select the preferred video in each pair. Since the comparison is only between the videos with and without enhancement, 3-level rating is used here: A is preferred to B, A is the same as B, and B is preferred to A. Subjects were also requested to report flickers in the videos if they saw any.

Eight subjects conducted the comparison experiment. No flicker was reported by any subjects. Therefore, we can claim the temporal consistency of the proposed content dependent enhancement method. The DMOS is computed between the enhanced videos and the original videos. Positive (or negative) numbers mean the enhanced (or original) videos are favored. The DMOS and the 95% CIs are plotted in Fig. 18. Note that the 3-level rating is performed here, so the opinion scores are among $\{-1, 0, 1\}$, instead of $\{-2, -1, 0, 1, 2\}$ used in the image comparison. As a result, the values of DMOS are about half of that in Fig. 17. Since fewer subjects are included in the video evaluation, the confidence intervals are larger than those of the image comparison. Generally, the results are consistent with the results of the image comparison. The enhanced videos are preferred to the original videos, and the preference is more obvious under brighter ambient light for a given L_W.

3) Further Discussion: The size of the display is expected to have subtle impact on the preference of images/videos. It is possible that subjects would pay less attention to small details when the images/videos are displayed on a smaller display. Detail loss caused by clipping the codewords at both ends might be neglected. Therefore, a brighter image with higher contrast at the mid-tone could be more favored.

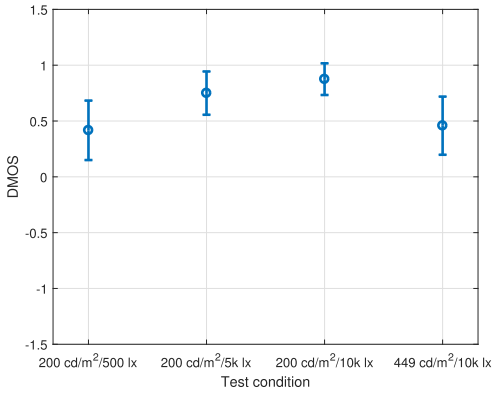


Fig. 18. 95% confidence intervals of average DMOS of proposed content dependent method vs. no processing for all tested videos. "5k" stands for 5000, and "10k" stands for 10,000.

The DMOS between the proposed methods and no processing drops slightly when L_W increases from 200 to 449 cd/m^2 which is the maximum L_W of the device. If the device could provide even higher L_W, it is possible that the advantage of the proposed methods could further decrease. However, as long as the screen reflects the ambient light, the contrast of dark codewords will be degraded, and thus the details in the dark regions of the original images/videos will be less visible. The proposed methods can improve the visibility of those regions, so the enhanced images/videos will still be favored.

B. Objective Comparison

In addition to the subjective tests, we compare different methods using objective metrics including absolute mean brightness error (AMBE) [26], discrete entropy (DE) [27], and measure of enhancement (EME) [28], to show luminance and contrast improvement on the enhanced images. The results are in Table I. The tone mapping curves generated by the method of Mantiuk et al. are steeper than the two proposed methods. Dark codewords are mapped to brighter values, and more codewords on the right end are clipped. As a result, AMBE and EME of the method of Mantiuk et al. are higher than those of the proposed methods, and DE of Mantiuk's method is lower. However, due to the loss of details, the images enhanced by Mantiuk's method are less preferred subjectively than the proposed methods. The tone mapping curves by the method of Su et al. are slightly less steep than the proposed methods, so its AMBE and EME are lower than the proposed methods. These simple objective metrics show only some aspects of the enhancement performance, and thus are not good predictors of the overall subjective quality.

We also compare different methods using dynamic range independent image quality assessment (DRI) [29]. Green color in Fig. 19 shows the loss of visible contrast, i.e., contrast is visible in the original image but invisible in the enhanced image. Red color indicates the amplification of invisible contrast, i.e., contrast is invisible in the original image but becomes visible in the enhanced image. The saturation of each color indicates the magnitude of detection probability. The method of Mantiuk et al. amplifies the contrast of dark regions the

TABLE I
AMBE, DE AND EME OF DIFFERENT METHODS

Metric	AMBE (cd/m^2)				DE				EME			
L_W	200			449	200			449	200			449
E_{amb}	500	5000	10,000	10,000	500	5000	10,000	10,000	500	5000	10,000	10,000
Original	-				6.99				14.26			
Mantiuk et al.	42.93	71.07	76.74	198.4	6.343	5.041	4.360	4.784	50.35	69.01	76.06	50.47
Su et al.	3.655	11.80	34.20	21.37	6.926	6.907	6.725	6.920	15.71	10.92	7.195	11.45
Proposed T^G	12.82	35.25	40.37	75.43	6.976	6.846	6.805	6.852	17.34	14.23	13.72	14.37
Proposed T^D	31.58	55.30	60.60	120.6	6.936	6.832	6.800	6.843	16.47	12.51	11.87	12.70

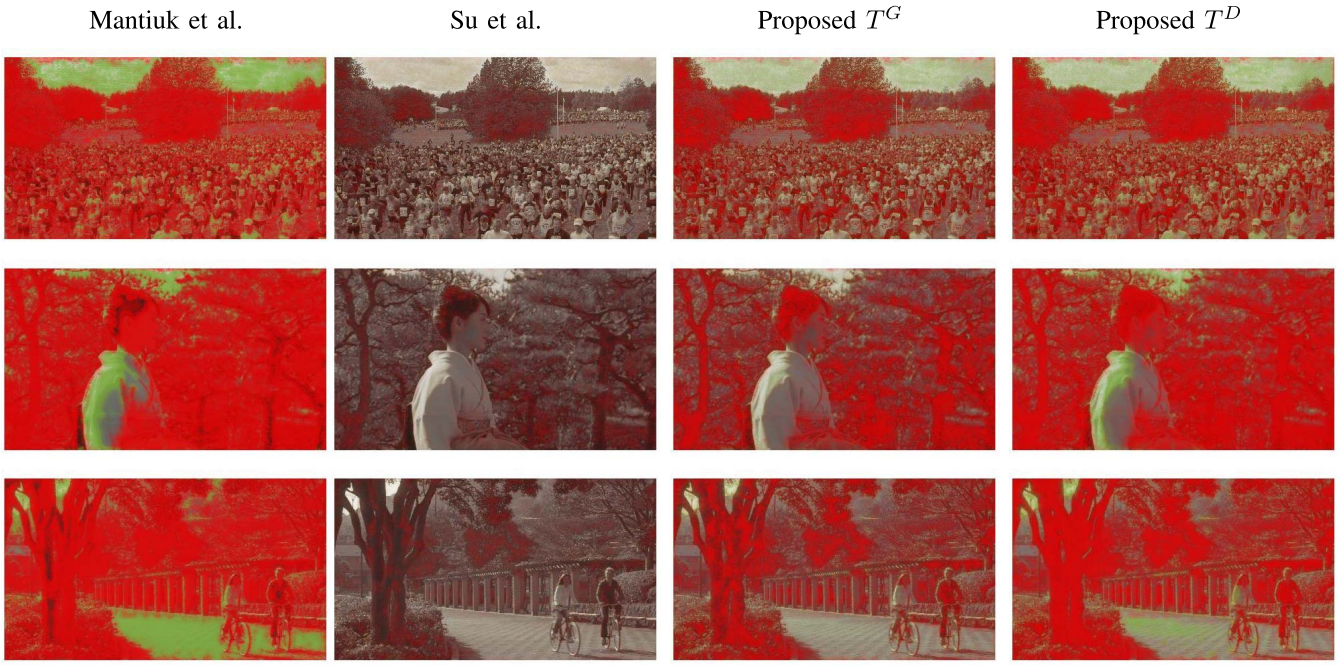


Fig. 19. DRI between original images and enhanced images using different algorithms for $L_W = 200$ and $E_{amb} = 5000$. Green stands for loss of visible contrast, and red for amplification of invisible contrast. The saturation of each color indicates the magnitude of detection probability.

most, but causes the most contrast loss in bright regions. The method of Su et al. has the least effect of enhancement among the four methods. The two proposed methods maintain good trade-offs between contrast amplification of dark regions and contrast loss of other regions.

C. Computational Times

The implementation in C on 3.3 GHz CPU (single-threading) of the proposed content independent tone mapping requires only 11.8 ms to build the mapping look-up tables (LUTs) for both luma and chroma channels. Because of the additional time required for histogram construction, the proposed content dependent tone mapping needs 22.3 ms to build the mapping LUTs for luma and chroma channels for a frame of high definition (HD) 1920×1080.

For an HD video frame, the average time to apply the LUTs sequentially to all the pixels is 12.0 ms. For a given ambient illumination, the speed to process an HD frame using the proposed content independent method is 83 frames per second (fps). If the ambient illumination happens to change at each frame, the overall processing speed is 42 fps. The speed of the proposed content dependent method to process an HD frame is 29 fps. Note that the processing time can be greatly reduced if the LUTs are applied to pixels in parallel. The computing time to construct histograms can be also reduced if parallel computing is employed.

V. CONCLUSION

We proposed two tone mapping operators to enhance the luminance and details of images and videos shown in bright ambient illumination. The tone mapping considers display characteristics and human visual sensitivity. The contrast loss in dark areas of images/videos due to reflected light and reduced sensitivity of eyes is compensated. The content independent tone mapping operator is constructed only once for the given viewing condition and can be applied to any content. The content dependent method uses simple statistics of content, and slightly outperforms the content independent method. We compared with the method of Mantiuk *et al.* [9] and the method of Su *et al.* [14], and our methods are preferred subjectively. Our proposed methods boost the visibility of details in dark areas and well preserve details in bright areas. The methods can be extended to high dynamic range videos which are represented by high bit depth. More accurate modeling of reflection would be interesting to explore in the future.

ACKNOWLEDGMENT

The authors would like to thank Dr. Louis J. Kerofsky who helped with display measurement.

REFERENCES

[1] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper*. Accessed: Feb. 6, 2017. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[2] J. H. Krantz, L. D. Silverstein, and Y.-Y. Yeh, "Visibility of transmissive liquid crystal displays under dynamic lighting conditions," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 34, no. 5, pp. 615–632, Oct. 1992.

[3] M. E. Miller and J. R. Niederbaumer, "Automatic luminance and contrast adjustment as functions of ambient/surround luminance for display device," U.S. Patent 6 529 212 B2, Mar. 4, 2003.

[4] *Display Mate website*. Accessed: Sep. 3, 2015. [Online]. Available: http://www.displaymate.com

[5] G. Nur, H. K. Arachchi, S. Dogan, and A. M. Kondoz, "Ambient illumination as a context for video bit rate adaptation decision taking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 12, pp. 1887–1891, Dec. 2010.

[6] G. Nur, H. K. Arachchi, S. Dogan, and A. M. Kondoz, "Extended VQM model for predicting 3D video quality considering ambient illumination context," in *Proc. 3DTV Conf. True Vis.-Capture, Transmiss. Display 3D Video (3DTV-CON)*, May 2011, pp. 1–4.

[7] F. C. Nur and G. Nur, "Prediction of 3D video experience from video quality and depth perception considering ambient illumination context," in *Proc. 1st Int. Conf. Future Gener. Commun. Technol.*, Dec. 2012, pp. 28–31.

[8] J. Xue and C. W. Chen, "Mobile video perception: New insights and adaptation strategies," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 3, pp. 390–401, Jun. 2014.

[9] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," in *Proc. ACM SIGGRAPH Papers (SIGGRAPH)*, 2008, Art. no. 68.

[10] H. Kobiki and M. Baba, "Preserving perceived brightness of displayed image over different illumination conditions," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 2485–2488.

[11] M.-Y. Lee, C.-H. Son, J.-M. Kim, C.-H. Lee, and Y.-H. Ha, "Illumination-level adaptive color reproduction method with lightness adaptation and flare compensation for mobile display," *J. Imag. Sci. Technol.*, vol. 51, no. 1, pp. 44–52, Jan./Feb. 2007.

[12] Y. J. Kim, "An automatic image enhancement method adaptive to the surround luminance variation for small sized mobile transmissive LCD," *IEEE Trans. Consum. Electron.*, vol. 56, no. 3, pp. 1161–1166, Aug. 2010.

[13] X. Xu and L. Kerofsky, "Improving content visibility for high-ambient-illumination viewable display and energy-saving display," *J. Soc. Inf. Display*, vol. 19, no. 9, pp. 645–654, 2011.

[14] H. Su, C. Jung, S. Wang, and Y. Du, "Adaptive enhancement of luminance and details in images under ambient light," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 1219–1223.

[15] H. Su, C. Jung, S. Wang, and Y. Du, "Readability enhancement of displayed images under ambient light," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2017.2676881

[16] *Reference Electro-Optical Transfer Function for Flat Panel Displays Used in HDTV Studio Production*, document ITU-R BT.1886, Mar. 2011.

[17] P. G. J. Barten, "Contrast sensitivity of the human eye and its effects on image quality," *Proc. SPIE*, vol. PM72, p. 232, Dec. 1999.

[18] *The Present State of Ultra High Definition Television*, document ITU-R BT.2246, Oct. 2011.

[19] P. G. J. Barten, "Formula for the contrast sensitivity of the human eye," *Proc. SPIE*, vol. 5294, pp. 231–238, Dec. 2003.

[20] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual signal coding for more efficient usage of bit codes," in *Proc. Annu. Techn. Conf. Exhib.*, Oct. 2012, pp. 1–9.

[21] *The Video Sequences Are From*. Accessed: Dec. 20, 2016. [Online]. Available: http://5.co.il/kimono1-and-park-scene-test-results

[22] *The Video Sequences Are From*. Accessed: Dec. 20, 2016. [Online]. Available: https://media.xiph.org/video/derf

[23] *Subjective Video Quality Assessment Methods for Multimedia Applications*, document ITU-T P910, Apr. 2008.

[24] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R BT.500-13, Jan. 2012.

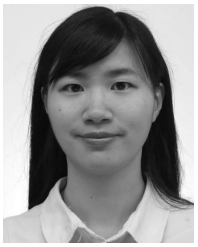
[25] R. Likert, "A technique for the measurement of attitudes," *Arch. Psychol.*, vol. 22, no. 55, p. 140, 1932.

[26] S.-D. Chen and A. R. Ramli, "Minimum mean brightness error bi-histogram equalization in contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 49, no. 4, pp. 1310–1319, Nov. 2003.

[27] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 2001.

[28] S. S. Agaian, B. Silver, and K. A. Panetta, "Transform coefficient histogram-based image enhancement algorithms using contrast entropy," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 741–758, Mar. 2007.

[29] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 69:1–69:10, Aug. 2008.



Qing Song (S'15–M'18) received the B.Eng. degree in automation from Tongji University, Shanghai, China, in 2011, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at San Diego, San Diego, CA, USA, in 2013 and 2017, respectively. She joined Dolby Laboratories, Inc., Sunnyvale, CA, USA, in 2017. Her research interests include image/video processing, compression, and transmission.

Pamela C. Cosman (S'88–M'93–SM'00–F'08) received the B.S. degree (Hons.) in electrical engineering from the California Institute of Technology in 1987 and the Ph.D. degree in electrical engineering from Stanford University in 1993. She held a post-doctoral position at the University of Minnesota. In 1995, she joined the Department of Electrical and Computer Engineering, University of California at San Diego, La Jolla, CA, USA, where she is currently a Professor. Her past administrative positions include the Director of the Center for Wireless Communications from 2006 to 2008, the ECE Department Vice Chair from 2011 to 2014, and the Associate Dean for students from 2013 to 2016. Her research interests are in image and video compression and processing, and wireless communications. Her awards include the ECE Departmental Graduate Teaching Award, the NSF Career Award, the Powell Faculty Fellowship, the Globecom 2008 Best Paper Award, the HISB 2012 Best Poster Award, the 2016 UC San Diego Affirmative Action and Diversity Award, and the 2017 Athena Pinnacle Award. She was an Associate Editor of the IEEE COMMUNICATIONS LETTERS and the IEEE SIGNAL PROCESSING LETTERS, and the Editor-in-Chief from 2006 to 2009. She was a Senior Editor of the IEEE *Journal on Selected Areas in Communications*.