

Video Restoration against Yin-Yang Phasing *

Xiaolin Wu^{1,2} Zhenhao Li¹ Xiaowei Deng¹

¹McMaster University, Canada

²Shanghai Jiao Tong University, China

xwu@mail.ece.mcmaster.ca {liz87,dengx23}@mcmaster.ca

Abstract

A common video degradation problem, which is largely untreated in literature, is what we call Yin-Yang Phasing (YYP). YYP is characterized by involuntary, dramatic flip-flop in the intensity and possibly chromaticity of an object as the video plays. Such temporal artifacts occur under ill illumination conditions and are triggered by object or/and camera motions, which mislead the settings of camera's auto-exposure and white point. In this paper, we investigate the problem and propose a video restoration technique to suppress YYP artifacts and retain temporal consistency of objects appearance via inter-frame, spatially-adaptive, optimal tone mapping. The video quality can be further improved by a novel image enhancer designed in Weber's perception principle and by exploiting the second-order statistics of the scene. Experimental results are encouraging, pointing to an effective, practical solution for a common but surprisingly understudied problem.

1. Introduction

1.1. The problem and background

A highly irritating type of video degradation, called Yin-Yang Phasing (YYP) in this paper, is spatially patchy and temporally inconsistent objects appearance, with parts of the scene turning in and out the state of under-exposure (yin) or over-exposure (yang), back and forth. Such effects are quite common in impromptu-made video materials in daily life. An example is presented in Figure 1: in a short progression of few video frames, which are produced by an iPhone 6 camera, the person's face alters from normally lit to unintelligibly dark as she moves her head. Moreover, the involuntary, dramatic changes in the intensity and possibly chromaticity are compounded by low contrast due to wrongly exposed objects. The YYP effects are caused by

unevenness of illumination, significant changes in the directions of incident and reflection lights due to object and/or camera motions, and improper camera operations by amateur users. These scenarios tend to confuse the ubiquitous "dumb" cameras and lead to incorrect and time-lagged settings of auto-exposure and white point. Unfortunately, the above root cause of YYP is an inherent weakness of mass-produced cameras that will trouble average users for a foreseeable future, particularly when shooting videos in unevenly-lit scenes.

In difficult illumination conditions, disabling camera's auto-exposure functionality cannot cure but rather aggravate the YYP problem for most users. For example, in backlit scenes like in Figure 1, without auto-exposure the person in the foreground, the very focus of the video session, will be severely underexposed in most frames. Also, out of question is the use of auxiliary lighting to compensate for problematic illumination in the scene, as it is beyond the means and knowledge of amateur camera users.

The YYP type of video degradation is becoming a major culprit of poor video quality, much more so than insufficient spatial resolution, low frame rate, sensor noises, and compression distortions. Nowadays even consumer-grade cameras boast very high pixel counts, frame rates of 60Hz and above, and low sensor noise level; furthermore, modern communication infrastructures can support high throughput visual data exchanges. While these hardware advances are making superresolution, frame rate upconversion, denoising, compression artifacts removal, etc., less important in practice, they can do nothing to repair the YYP degradation as the lighting conditions in the scene and non-expert video shooting behaviors are beyond the control of video acquisition and communication equipment. Despite the daily encountering of the YYP video degradation problem in a wide range of video applications, such as social media, on-line video sharing, video monitoring, spontaneous video reporting, etc., very little research has been carried out on the YYP phenomenon and its mitigation. In this paper we investigate how to algorithmically remove the undesired YYP effects, aiming to restore the video to the state as though it

*This work was supported by Natural Sciences and Engineering Research Council of Canada, and Natural Science Foundation of China, grant NSFC 61331014.



Figure 1. Top row: temporally inconsistent frames due to changes in exposure time. Bottom row: frames restored by the proposed method.

was shot in spatially uniform and temporally steady lighting conditions. For lack of better wording, we call such a process YYP removal.

Unlike in the studies of traditional image/video restoration problems, such as deshaking [10], denoising [9], deblurring [2], superresolution [12], etc., the source of YYP degradation is not purely physical and hence difficult to model analytically. As a result, casting YYP removal as an inverse problem or into a mathematical programming formulation is not as easy as for other restoration problems.

1.2. Our approach

As tonal reproduction primarily depends on global image statistics, it is difficult to perform the YYP removal solely in pixel domain by tracking motion trajectory across frames and enforcing temporal tone consistency guided by the motion flow. Instead, we formulate the YYP removal as an inter-frame, spatially-adaptive, optimal tone mapping problem and propose an optimization approach to solve it, aiming to neutralize temporally unsteady intensity levels of the objects. The basic premise of our approach is that the overall appearance of an image is governed by the shape of its intensity histogram [11]. Therefore, we propose to retain the temporal consistency of tonal reproduction by mapping the intensity histograms of input frames to a common target histogram per a given scene. The other reason for adopting a histogram-based tone mapping approach is that it can be tuned to boost contrast as well, hence unifying the tasks of global stabilization and image enhancement.

As the human visual system can rapidly adapt to differ-

ent luminance levels, our histogram targeting strategy for YYP harmonization can be made more effective if the tone mapping is tailored to image regions under different illumination conditions. Specifically, we segment a YYP-affected video frame into two types of regions, by the likelihood of a pixel being on a weakly or strongly illuminated object surface, denoted by the W -region and S -region. The segmentation is performed via homomorphic filtering and a fuzzy classification.

In order to prevent the W -regions from being underexposed and the S -regions from being overexposed sporadically, we temporally track the W -regions and S -regions. For a fixed video scene of N frames, let $R_W^{(n)}$ and $R_S^{(n)}$, $1 \leq n \leq N$, be the sets of W -regions and S -regions in frame n , respectively. Using the first-order and second-order statistics of the data set $\{R_W^{(n)}\}_{n=1}^N$, an anchor intensity histogram \mathbf{h}_W^* is constructed as the target for the YYP harmonization of W -regions through all N frames, which can be viewed as a generalized centroid in terms of Kullback-Leibler distance of probability distributions [3]. Then, for each frame n , a tone mapping (histogram transformation) $T_W^{(n)}$ is computed to best match the histogram of $R_W^{(n)}$ to \mathbf{h}_W^* . By forcing all N histogram transformations $T_W^{(n)}$, $1 \leq n \leq N$, to approach the same output histogram \mathbf{h}_W^* , we make the tone reproduction of an object temporally consistent. In addition, video enhancement can be incorporated into the proposed YYP harmonization framework. The anchor histogram \mathbf{h}_W^* can be designed to maximize a Lagrangian of a Weber-law contrast metric and the entropy

of \mathbf{h}_W^* . This optimization problem has an efficient dynamic programming solution. The same procedure outlined above can be applied to the set of S -regions $\{R_S^{(n)}\}_{n=1}^N$ to compute \mathbf{h}_S^* and $T_S^{(n)}$, $1 \leq n \leq N$. The main objective of the proposed YYP harmonization algorithm is to stabilize throbbing intensity levels in any given W - or S -region. But it can also stabilize shifting chromaticity of any given region, if required. We only need to first decompose the original video signal into luminance and chrominance components, and then apply the same region-adaptive inter-frame harmonization method to the chrominance component of the input video.

To summarize our novel YYP harmonization approach, objects under similar lighting conditions are spatially grouped and temporally tracked, the intensity and chrominance distributions of these similarly-lit objects are optimized in terms of perceptual quality, and set as the objectives of tone mapping for all frames in a given scene of the YYP-degraded video. For each frame n , to prevent possible boundary effects of region-based tone mapping functions $T_W^{(n)}$ and $T_S^{(n)}$, the results of $T_W^{(n)}$ and $T_S^{(n)}$ are weighted based on a fuzzy light field segmentation.

1.3. Related works

Thus far most of research efforts on video restoration against adverse acquisition conditions are devoted to the stabilization of shaky frames caused by large, irregular camera jitters [10]. Of most relevance to this work is a 2011 publication by Farbman and Lischinski on tonal stabilization of video [6]. The authors addressed the problem of tonal inconsistency in consecutive video frames caused by improper auto-exposure and white points, and proposed a method to make global, smooth tone transitions from one frame to the next guided by selected anchor frames. In comparison, YYP is a different and more challenging video degradation problem: severe tonal fluctuations *compounded* by low contrast, which are caused by uneven, incorrect exposures. In [6] two assumptions are made: 1. lighting conditions in the scene do not change abruptly; 2. the tonal fluctuations are of a global nature not spatially varying; moreover, there is no significant loss of contrast due to under and/or over exposures. The YYP problem differs from the one in [6] in both the effect and the cause. YYP removal requires an approach of joint tone stabilization and contrast enhancement, as proposed by this paper.

The YYP phenomenon has similar visual characteristics as poor tone reproduction in high dynamic range images (HDR). If each video frame is treated in isolation, it is tempting to try some of the many HDR tone mapping methods [4, 13] to correct poor use of dynamic range and improve the visual quality. However, this naive approach is highly prone to objectionable temporal artifacts. Very recently, Aydin *et al.* addressed the problem of temporal co-

herence in tone mapping of HDR video [1]. They proposed a method of edge-aware filtering method through pixel motion paths to achieve temporal stability of the enhanced video. But this work, like other HDR tone mapping methods, is mainly about how to compress the intensity dynamic range while maintaining contrast and preventing artifacts; hence it is not suited to compensate for drastic changes in object appearance in time as required by the restoration of YYP videos.

The remainder of this paper is structured as follows. Section 2 presents an algorithm for temporally-constrained fuzzy light field segmentation, which is a preparation step for the main task of inter-frame region-adaptive YYP harmonization. Section 3 details the YYP harmonization algorithm, in particular explaining the choice of an anchor frame and the histogram targeting process. Section 4 discusses how to improve the performance of YYP harmonization by using an enhanced anchor frame to drive the algorithm developed in Section 3. Section 5 reports experimental results and performance evaluations.

2. Temporally-constrained fuzzy light field segmentation

As reasoned in the introduction, our YYP harmonization method needs to analyze an unevenly-illuminated video scene, and separate weakly-illuminated W -regions from strongly-illuminated S -regions in the input frame. To this end, the input video signal is decomposed into the luminance and chrominance components. For the purpose of segmenting the luminance image $I(x, y)$ by surface illumination strength, we adopt the following image formation model [7]

$$I(x, y) = L(x, y) \cdot R(x, y) \quad (1)$$

where $L(x, y)$ is the light energy striking on the surface position corresponding to pixel (x, y) , and $R(x, y)$ is the reflectance of the surface point. Even with great discrepancy in surface illumination in the YYP phenomenon the light field function $L(x, y)$ is still piecewise smooth, or being a low-pass signal; in comparison, the reflectance signal $R(x, y)$ consists of higher frequency components than $L(x, y)$. Therefore, we can apply the non-linear homomorphic filtering to extract the 2D surface illumination function $L(x, y)$ from the input image $I(x, y)$. The next step is to detect and describe W -regions and S -regions in the low-pass illumination image $L(x, y)$. By observing that the illumination strength on a given object surface is nearly constant, we model the 2D illumination function $L(x, y)$ to be piecewise constant. The split-and-merge segmentation algorithm based on piecewise constant approximation [15] is suited for our segmentation task and applied to each frame n ; the resulting segments are then classified as W - or S -regions by the least-squares thresholding, and placed into set $\mathcal{R}_W^{(n)}$

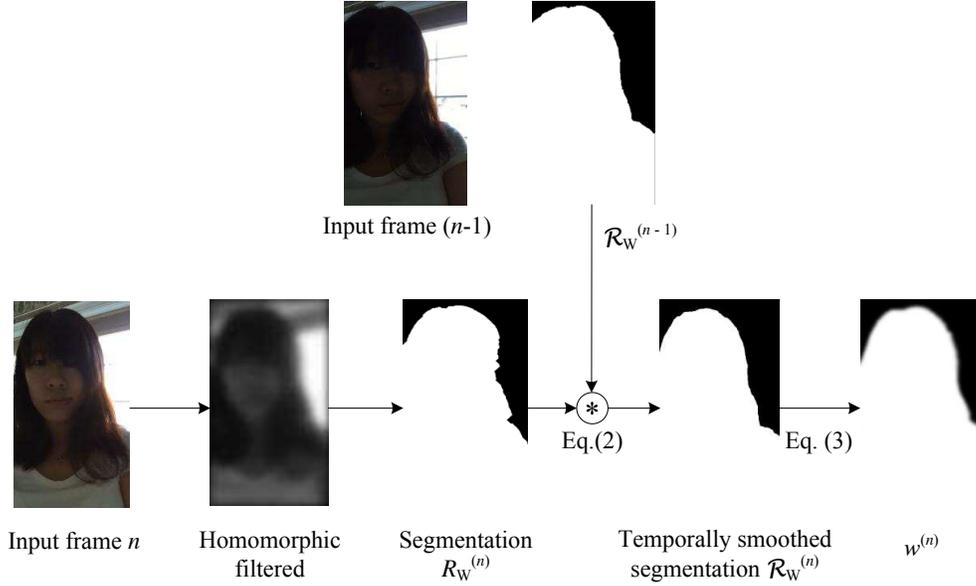


Figure 2. Steps of temporally-constrained fuzzy segmentation.

or set $\mathcal{R}_S^{(n)}$ as defined in the introduction. In order to prevent the formation of too small W -regions and S -regions, a lower bound on the segment size should be set in the split-and-merge segmentation. As the reader will appreciate shortly when we present the details of the YYP harmonization algorithm, the precision of region boundaries is less critical for the algorithm performance. However, the above proposed segmentation process can be refined, if desired, by one of many edge-aware filtering techniques that can utilize the color information as well [8]. For visual quality of video restoration, more important than boundary precision is that the segmentation results should be consistent in time. Under the assumption that in a given video scene, the foreground object motion or/and camera motion is modest, we can achieve the desired temporal consistency by the following simple and fast inter-frame segmentation technique:

$$\mathcal{R}_j^{(n)} = ((\mathcal{R}_j^{(n-1)} \cap R_j^{(n)}) \oplus \tau) \cap (\mathcal{R}_j^{(n-1)} \cup R_j^{(n)}), \quad (2)$$

$$j \in \{W, S\}$$

where τ is the core for morphologic dilation, $\mathcal{R}_W^{(n)}$ ($\mathcal{R}_S^{(n)}$) is the temporally smoothed $R_W^{(n)}$ ($R_S^{(n)}$) with respect to its counterpart $\mathcal{R}_W^{(n-1)}$ ($\mathcal{R}_S^{(n-1)}$) in the previous frame. Set $\mathcal{R}_W^{(n)}$ (or reciprocally set $\mathcal{R}_S^{(n)}$) deterministically classifies every pixel in frame n to be on a weakly-illuminated (or strongly-illuminated) object surface or not. Such a hard-decision classification may, due to segmentation errors, generate boundary artifacts after pixels in $\mathcal{R}_W^{(n)}$ and $\mathcal{R}_S^{(n)}$ are gone through two different tone mappings $T_W^{(n)}$ and

$T_S^{(n)}$. We introduce a simple fuzzy classification technique to eliminate the boundary artifacts. The idea is to weigh the results of $T_W^{(n)}$ and $T_S^{(n)}$ by the likelihood of a pixel being on a weakly-illuminated object surface. A fuzzy classifier $w^{(n)}$ is generated by convoluting the hard segmentation $\mathcal{R}_W^{(n)}$ (which takes 1 at W -region and 0 at S -region) of 2D image with a Gaussian kernel g :

$$w^{(n)} = \mathcal{R}_W^{(n)} * g \quad (3)$$

The value of $w^{(n)}(x, y)$ will be used as the likelihood for pixel (x, y) being on weakly-illuminated object surface. Figure 2 summarizes the temporally-constrained fuzzy segmentation-classification process presented in this section.

3. Interframe region-adaptive YYP harmonization

Now we discuss how to construct the tone mappings $T_W^{(n)}$ and $T_S^{(n)}$ for restoring the interframe consistency of frame n in a YYP-degraded video. All of the following technical developments, conclusions and methods apply exactly the same way to the construction of $T_W^{(n)}$ and $T_S^{(n)}$. Therefore, we can drop the subscripts in previous notations T_W , $\mathcal{R}_W^{(n)}$, T_S , $\mathcal{R}_S^{(n)}$, etc. to avoid symbol clutter, and discuss the case of $T_W^{(n)}$ only.

Let $\mathbf{h}^{(1)}$, $\mathbf{h}^{(2)}$, \dots , $\mathbf{h}^{(N)}$ be the intensity histograms of W -regions for the N input frames in a given video scene. We need to select, among the N input frames, frame n^* such that $\mathbf{h}^{(n^*)}$ is statistically the best representative of all

other $N - 1$ histograms. Histogram $\mathbf{h}^{(n^*)}$ can be viewed as a generalized centroid of the set $\{\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(N)}\}$ and is used as the target intensity distribution for all tone-mapped frames by $T^{(n)}$, $1 \leq n \leq N$, to closely obey.

Using the Kullback-Leibler distance $D(\cdot||\cdot)$ between distributions, the generalized centroid histogram $\mathbf{h}^{(n^*)}$ can be computed as below:

$$\begin{aligned} \mathbf{h}^{(n^*)} &= \arg \min_{\mathbf{h} \in \{\mathbf{h}^{(n)}\}_{n=1}^N} \sum_{j=0}^N D(\mathbf{h}^{(j)}||\mathbf{h}) = \arg \min_{\mathbf{h}} \sum_{j=0}^N H(\mathbf{h}^{(j)}, \mathbf{h}) \\ &= \arg \min_{\mathbf{h}} H(\bar{\mathbf{h}}, \mathbf{h}) \end{aligned} \quad (4)$$

In other words, among all N frames of the scene, frame n^* is the one whose histogram has the minimum cross entropy with respect to the average histogram $\bar{\mathbf{h}}$ of the group.

Upon having selected the anchor intensity histogram $\mathbf{h}^{(n^*)}$, the tone mapping function $T^{(n)}$ for W -regions of frame n , $1 \leq n \leq N$, is computed via histogram matching:

$$T^{(n)}(k) = \arg \min_j |C^{(n)}(k) - C^{(n^*)}(j)| \quad (5)$$

where C stands for the cumulative density function of the corresponding histograms.

The next step is to eliminate possible boundary artifacts caused by separate tone mappings $T_W^{(n)}$ and $T_S^{(n)}$ on W - and S -regions, respectively, as explained in the previous section. Recall that the proposed fuzzy segmentation assigns each pixel (x, y) a likelihood value $w(x, y) \in [0, 1]$, with the pixels around the border of W - and S -regions being far off from 0 and 1. In the interest of robustness, the input intensity value $I(x, y)$ is finally mapped to $\tilde{I}(x, y)$ via the following affine weighting by the region likelihood $w(x, y)$ of the pixel:

$$\tilde{I}(x, y) = w(x, y)T_W^{(n)}(I(x, y)) + (1 - w(x, y))T_S^{(n)}(I(x, y)) \quad (6)$$

In the same approach to harmonizing time-varying luminance in YYP-degraded video, we can neutralize temporal variations of chromaticity, if necessary, by matching the chrominance distributions of all frames to that of the anchor frame n^* . Any two-dimensional chrominance space, such as (U, V) in YUV or (H, S) in HSI , can be used; in this paper we adopt the (U, V) chrominance space. In order for the (U, V) distribution of frame n with mean vector $\mu^{(n)}$ and covariance matrix $\Sigma^{(n)}$ to match that of the anchor frame n^* with mean vector μ^* and covariance matrix Σ^* , we solve the following optimization problem

$$\begin{aligned} \{\mathbf{A}^{(n)}, \mathbf{t}^{(n)}\} &= \\ \arg \min_{\mathbf{A}, \mathbf{t}} &\|\mathbf{A}\mu^{(n)} + \mathbf{t} - \mu^*\|_2^2 + \|\mathbf{A}\Sigma^{(n)}\mathbf{A}^T - \Sigma^*\|_F^2. \end{aligned} \quad (7)$$

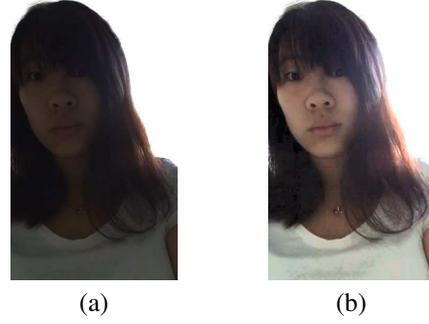


Figure 3. (a) a YYP-degraded frame n^* (W -region); (b) the enhanced version by the method proposed in section 4.

to determine the affine transform $\mathbf{A}^{(n)}$ and the translation vector $\mathbf{t}^{(n)}$. It can be shown (derivation details dropped to save space) that transforms $\mathbf{A}^{(n)}$ and $\mathbf{t}^{(n)}$, when applying to all pixel chrominance vectors (U, V) of frame n , achieve the best chrominance match between anchor frame n^* and input frame n in a least-squares sense.

4. Contrast enhancement in YYP restoration

Often after the YYP degradation, the representative frame n^* chosen by (4) is still of poor visual quality, with ill-shaped histograms $\mathbf{h}_W^{(n^*)}$ and $\mathbf{h}_S^{(n^*)}$. In such cases we can and should first enhance the anchor frame n^* , and then drive the YYP harmonization algorithm with the histograms of the enhanced anchor frame.

Any contrast enhancement can be applied to generate an improved anchor frame and the corresponding histograms of W - and S -regions. In this section, we propose an enhancement algorithm based on the second-order statistics of the input frame n^* and Weber's perception law. The effect of anchor frame enhancement is shown in Fig. 4.

According to Weber's perception law [7], a metric C for image contrast is

$$C \propto \frac{\Delta I}{I} \quad (8)$$

which is the ratio of local signal variation ΔI over the signal (stimulus) strength I . Based on Weber's perception principle and the second-order statistics of the input image, we define the expected contrast $C(T)$ for a histogram transform-based tone mapping T to be

$$\begin{aligned} C(T) &= \sum_{i=0}^{L-1} \sum_{j=1}^{L-1} p_{ij} \left[\frac{T(j) - T(i)}{j + i} \right] \\ \Rightarrow C(\mathbf{s}) &= \sum_{i=0}^{L-1} \sum_{j=1}^{L-1} p_{ij} \cdot \frac{1}{j + i} \sum_{k=i+1}^j s_k \\ &= \sum_{k=1}^{L-1} s_k \sum_{i=0}^{k-1} \sum_{j=k}^{L-1} \frac{p_{ij}}{j + i} \end{aligned} \quad (9)$$

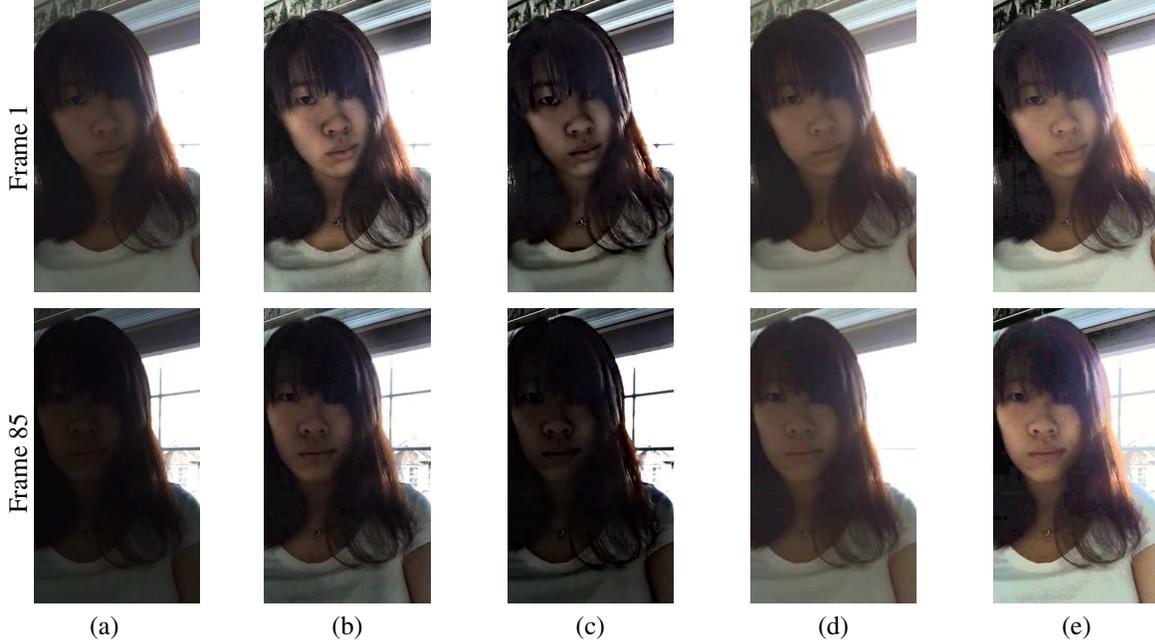


Figure 4. Rows: two YYP-degraded frames and corresponding restored results. Columns: (a) original; (b) output by CLAHE [16]; (c) output by method [5]; (d) output by method [6]; (e) output by the proposed method.

where p_{ij} is the joint probability of two spatially adjacent pixels taking on gray levels i and j respectively, $s_k = T(k) - T(k - 1)$, and L is the number of gray levels supported by the display device. In (9), $\frac{T(j)-T(i)}{j+i}$ is the local contrast of the pixels based on Weber's law (8). The expected contrast $C(T)$ can be expressed as a function of vector $\mathbf{s} = (s_1, s_2, \dots, s_L)$; moreover, the objective function $C(\mathbf{s})$ is linear in \mathbf{s} :

$$C(\mathbf{s}) = \varphi^T \mathbf{s}, \quad \varphi_k = \sum_{i=0}^{k-1} \sum_{j=k}^{L-1} \frac{p_{ij}}{j+i} \quad (10)$$

It should be noted that there is a one-to-one correspondence between T and \mathbf{s} under the constraint $\mathbf{s} \geq 0$; equivalently, \mathbf{s} can also represent histogram transform or tone mapping function.

Contrast enhancement calls for maximizing the expected contrast $C(\mathbf{s})$. But high visual quality also requires texture richness, which is associated with high entropy $H(\mathbf{h}^*(\mathbf{s}))$, $\mathbf{h}^*(\mathbf{s})$ being the histogram of anchor frame n^* after being enhanced by tone mapping \mathbf{s} . Finally, we determine the optimal tone mapping \mathbf{s} by solving the following optimization problem:

$$\begin{aligned} \mathbf{s}^* &= \arg \max_{\mathbf{s}} \{C(\mathbf{s}) + \lambda H(\mathbf{h}^*(\mathbf{s}))\} \\ \text{s.t.} \quad &\sum_{k=1}^{L-1} s_k = L, \quad \mathbf{s} \geq 0 \end{aligned} \quad (11)$$

where the Lagrangian multiplier λ is used to adjust the relative importance of edge sharpness and tone continuity. In

the above development, $\mathbf{h}^*(\mathbf{s})$ can be replaced by either $\mathbf{h}_W^*(\mathbf{s})$ and $\mathbf{h}_S^*(\mathbf{s})$. In other words, the same enhancement algorithm can be applied to improve $\mathbf{h}_W^{(n^*)}$ and $\mathbf{h}_S^{(n^*)}$ for the purpose of better YYP restoration as described at the opening of this section.

Algorithmically, the discrete optimization problem (11) is of the same structure as that of optimal scalar quantizer design; thus it can be solved efficiently by an optimal quantization algorithm [14] with a straightforward modification in the cost function.

The novelty of our image enhancement technique is to take expectation of a Weber's contrast metric $C(\mathbf{s})$ over the joint probability p_{ij} . The second-order statistics p_{ij} is used because the very notation of contrast directly relates to spatially adjacent pixel pairs.

5. Experimental results and performance evaluation

We conducted extensive experiments with the proposed YYP restoration method on videos captured by smartphones and laptops under poor, uneven illumination conditions and with both camera and object motions. Some samples of our experimental results are presented below; more example videos are available as supplementary materials on the internet.

Because the YYP-type of video degradation as identified by this paper has hardly been treated in the literature, there are no previous YYP video restoration methods to compare with. A technique of close spirit to ours is the one on video

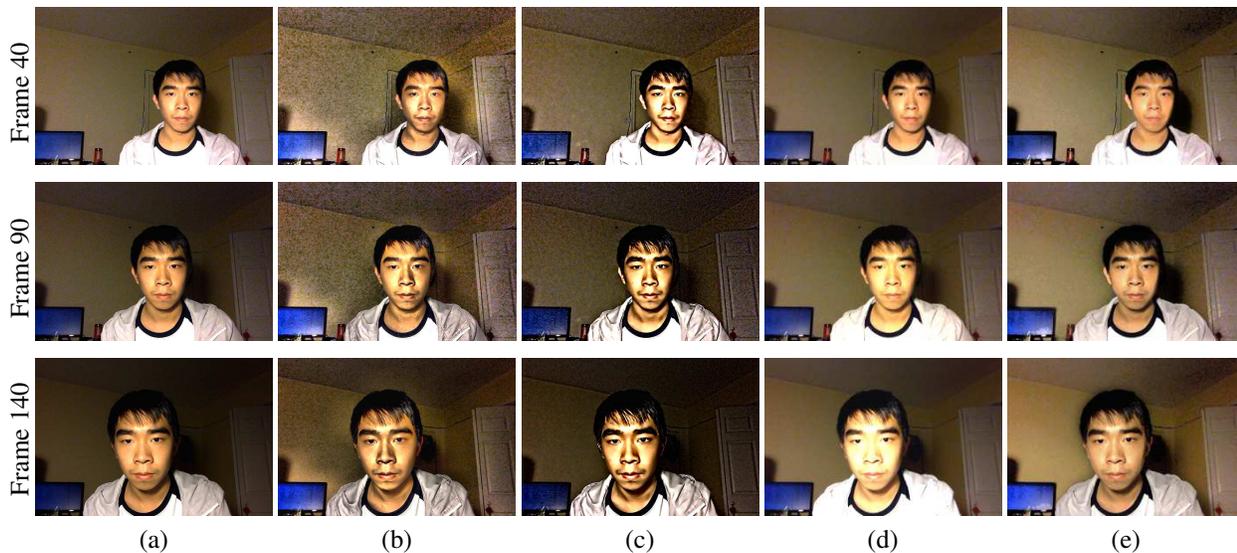


Figure 5. Rows: three YYP-degraded frames and corresponding restored results. Columns: (a) original; (b) output by CLAHE [16]; (c) output by method [5]; (d) output by method [6]; (e) output by the proposed method.

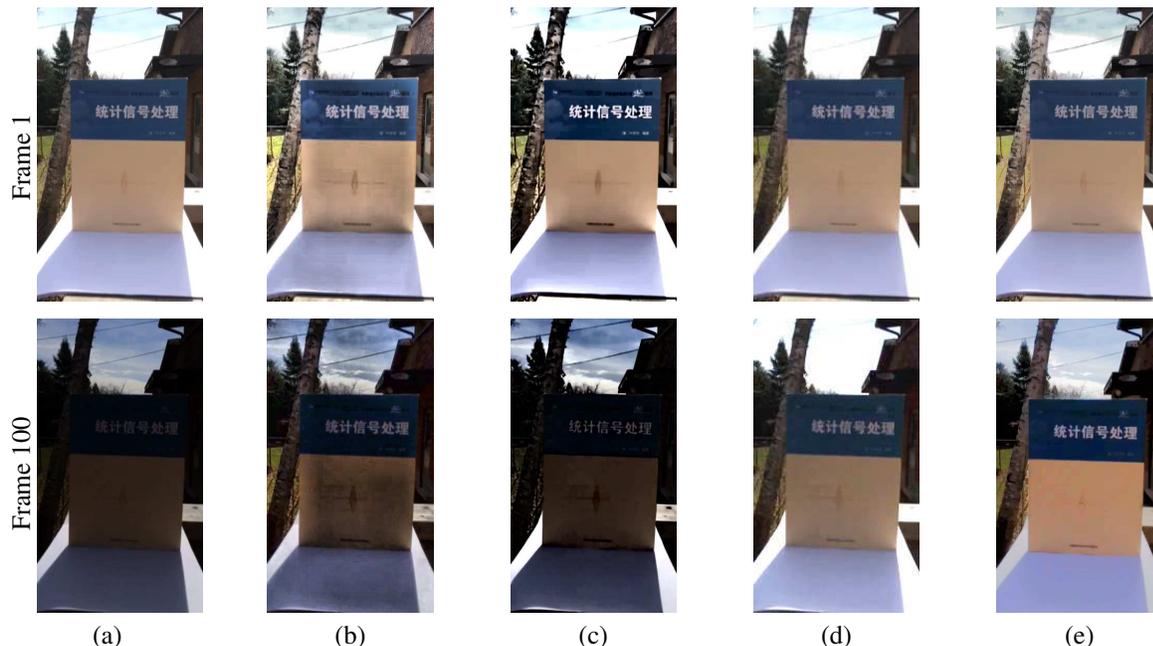


Figure 6. Rows: two YYP-degraded frames and corresponding restored results. Columns: (a) original; (b) output by CLAHE [16]; (c) output by method [5]; (d) output by method [6]; (e) output by the proposed method.

tonal stabilization by Farbman *et al.* [6]; this work is compared with the proposed YYP harmonization method. Since a simple way of attacking the YYP problem is to perform tone mapping on each input frame, we add to our comparison and evaluation group two single-frame methods: the contrast limited adaptive histogram equalization (CLAHE) [16], which represents the family of histogram transform-based global tone mapping methods; and the enhancement method based on edge-preserving decomposition (EPD) by

Farbman *et al.* [5], which represents the family of edge-aware local filter-based tone mapping methods. The video tonal stabilization method also requires an anchor frame; this anchor frame is chosen manually to be an input frame of best visual quality for fair comparison.

In Figure 4 we compare the above four methods and demonstrate how they behave when applied to a YYP-degraded header-and-shoulder video. This scenario of face-to-face video communication is very common in social me-

dia (e.g., Apple’s FaceTime), and it is highly susceptible to YYP degradation because the video is frequently shot in unfavorable indoor lighting, inexpensive cameras on mobile devices are limited in optical capability, and users tend to be naive in operating the camera.

The video scene in Figures 4 and 1 is backlit. The uneven illumination, as explained in the introduction, causes the camera’s auto-exposure to flip-flop accompanying the motions of the person or/and camera. The foreground person (W -region) becomes severely underexposed from time to time, meanwhile the background (S -region) is also unstable and suffers from overexposure intermittently (see the uploaded video supplementary materials). As shown in Figure 4 and the supplementary video file, the two single-frame methods are ineffective to neutralize the temporal intensity fluctuations, particularly on the face. The method [6] performs much better than the single-frame methods in terms of the temporal consistency in the foreground intensity, but it fails to correct the overexposure problem in the background. The proposed method appears to be more effective and robust than all others; it removes frame-to-frame intensity drifting in both foreground and background and at the same time enhances under- and over-exposed regions, greatly boosting the video quality.

The YYP degradation shown in 5 has a different polarity from that in Figure 4: the background is underlit and underexposed, where the person in the foreground is well lit. The background becomes even darker as the person moves closer to the light source (frame 170 in the figure). This generates serious halo artifacts in CLAHE output. The method [5] fails to bring any temporal consistency to the background intensity. The method [6] stabilizes the time-varying background intensity but in the process it leaves the foreground object overexposed. Again, the proposed method performs noticeably better than others in both W -region (the background) and S -region (the foreground object), retaining temporal consistency and rich spatial details.

Figure 6 shows a YYP-degraded video shot in an outdoor situation, together with restored results by the four different methods. Here the YYP phenomenon happens when the camera focus moves from the book (frame 1) to the sky (frame 100). The two single-frame methods fail to correct the underexposure problem in the foreground (frame 100) and they leave the drastic intensity changes in time largely uncompensated. The method [6], on the other hand, does a better job in the temporal consistency of overall intensity; but it makes the sky severely overexposed, wiping out details such as the clouds. In comparison, the proposed method effectively mitigates the underexposure problem in the W -region (the book) without overexposing the S -region (the sky); its restored video has a more steady tone reproduction largely immune to camera motion.

To further validate the efficacy of the proposed YYP har-

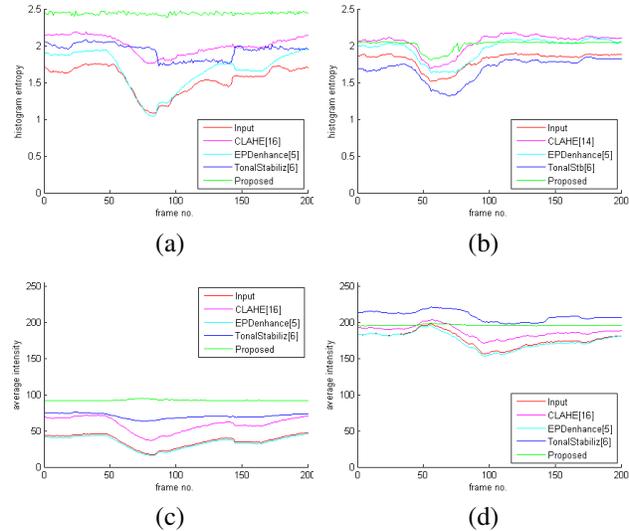


Figure 7. Performance comparison of different methods on video sequence of Fig. 1. (a) Entropy of the W -region; (b) Entropy of the S -region; (c) Average intensity of the W -region; (d) Average intensity of the S -region.

monization method, we evaluate it and its alternatives in two objective metrics as well. The first metric is the entropy of the restored video. The entropy can measure both the temporal consistency and the detail richness of the restored video. In Figure 7 we plot separately the entropies of W -regions and S -regions that are restored by the four different methods. The plotted curves step through all frames of the video scene in Figure 1 to demonstrate the temporal behaviors of the different methods. As being evident in the figure, the proposed method has the highest entropy in W -regions by a significant margin and almost ties for the highest entropy in S -regions, corroborating our empirical findings that the proposed method reproduces richer details than other methods. Moreover, the proposed method has nearly flat entropy curves, whereas other methods have much varied entropy values in time. This distinction is also clear by the second objective metric: the average frame luminance (Figure 7(c)(d)). The restored video by the proposed method keeps a nearly constant luminance in both W - and S -region, while the two single-frame methods suffer from intensity fluctuations, and the tonal stabilization method sits in between.

References

- [1] T. O. Aydin, N. Stefanoski, S. Croci, M. Gross, and A. Smolic. Temporally coherent local tone mapping of hdr video. *ACM Transactions on Graphics (TOG)*, 33(6):196, 2014.
- [2] S. Cho and S. Lee. Fast motion deblurring. In *ACM Transactions on Graphics (TOG)*, volume 28, page 145. ACM, 2009.

- [3] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.
- [4] G. Eilertsen, R. Wanat, R. K. Mantiuk, and J. Unger. Evaluation of tone mapping operators for hdr-video. In *Computer Graphics Forum*, volume 32, pages 275–284. Wiley Online Library, 2013.
- [5] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)*, volume 27, page 67. ACM, 2008.
- [6] Z. Farbman and D. Lischinski. Tonal stabilization of video. *ACM Transactions on Graphics (TOG)*, 30(4):89, 2011.
- [7] R. C. Gonzalez. *Digital image processing*. Pearson Education India, 2009.
- [8] K. He, J. Sun, and X. Tang. Guided image filtering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(6):1397–1409, 2013.
- [9] M. Mahmoudi and G. Sapiro. Fast image and video denoising via nonlocal means of similar neighborhoods. *Signal Processing Letters, IEEE*, 12(12):839–842, 2005.
- [10] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum. Full-frame video stabilization with motion inpainting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(7):1150–1163, 2006.
- [11] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson. Image statistics and the perception of surface qualities. *Nature*, 447(7141):206–209, 2007.
- [12] W.-C. Siu and K.-W. Hung. Review of image interpolation and super-resolution. In *Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pages 1–10. IEEE, 2012.
- [13] K. Smith, G. Krawczyk, K. Myszkowski, and H.-P. Seidel. Beyond tone mapping: Enhanced depiction of tone mapped hdr images. In *Computer Graphics Forum*, volume 25, pages 427–438. Wiley Online Library, 2006.
- [14] X. Wu. Optimal quantization by matrix searching. *Journal of algorithms*, 12(4):663–673, 1991.
- [15] X. Wu. Adaptive split-and-merge segmentation based on piecewise least-square approximation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(8):808–815, 1993.
- [16] K. Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics gems IV*, pages 474–485. Academic Press Professional, Inc., 1994.