Temporal Frequency of Flickering-Distortion Optimized Video Halftoning for Electronic Paper

Chao-Yung Hsu, Chun-Shien Lu, *Member, IEEE*, and Soo-Chang Pei, *Fellow, IEEE*

Abstract—Video halftoning is a key technology for use in electronic paper (e-paper) or smart paper, which is an emerging display device that has received considerable attention recently. In this paper, a temporal frequency of flickering-distortion optimized video halftoning method is proposed. We first uncover three visual defects that conventional neighboring frame referencing-based video halftoning methods, due to their sequential changes of reference frames, will encounter. To deal with the problem, we then propose a reference frame update per GOP-based error diffusion video halftoning method based on a flickering sensitivity-based human visual model. To efficiently compromise between average temporal frequency of flickering (ATFoF) and visual quality, temporal frequency of flickering-distortion (TFoFD) is presented as a metric for video halftoning performance evaluation. Based on the proposed probability model of video halftoning, the TFoFD curve can be accurately estimated to optimize the tradeoff between quality and ATFoF before the video is halftoned. Our temporal frequency of flickering-distortion optimization strategy can also be applied to other video halftoning schemes for performance improvement. Experimental results and comparisons with known methods demonstrate the effectiveness of our video halftoning method.

Index Terms—Electronic paper, flickering, flickering sensitivity, temporal frequency of flickering-distortion (TFoFD), video halftoning.

I. INTRODUCTION

A. Background

I N the film, "Minority Report," there is a clip showing John Anderton (acted by Tom Cruise) boarding a train to hide himself in the crowd in order to escape from his partner. Before he enters the train, he is captured and identified by a surveillance camera. Soon, the focus of the picture moves from John's face to the newspaper shown on an electronic device owned by a passenger who sits opposite John. As shown in the movie (from

Fig. 1. Digital halftoning process.

00:46:01 to 00:47:40),[1] we can see that some columns of the newspaper are changed to display some news about John. It was just a movie in 2002, but it is becoming reality now.

During the past few years, many companies have devoted themselves to manufacturing the first high-resolution electronic ink-based display module for reading-intensive applications. In particular, the first e-paper manufacturer[2] began to sell its microcapsule-based products in 2005. Specifically, its technique sees white particles suspended in a dark liquid, i.e., when an electric field is passed through them, the white particles get moved up or down. In 2006, a new product of e-paper[3] was announced at Hitachi uVALUE Convention 2006 that possesses the features of higher resolution (1024×768), larger screen size (3.1 in), and color display. Another recent e-paper product[4] demonstrates considerably greater capabilities than the earlier products. These facts indicate that the bit-depth of microcapsule-based e-paper is 1 b/color. The only exception (larger than 1 b/color) we can find is the color e-paper device[5], which possesses the features of color display, high resolution (768×1014), and large size (7.8 in).

By observing the developing status of e-paper, it may be envisioned that, in the future, most applications based on traditional paper will find an alternative such as e-paper. Since digital halftone images are the only format that microcapsule-based devices can display, it is reasonable to conjecture that halftone video data will be the next media that can be displayed on e-paper. In view of the fact that most e-paper devices developed so far are 1 bit-depth, the major goal of our study is to develop a new video halftoning method to meet the requirement of e-paper. To our knowledge, this is still a rather unexplored field.

---

[1]These video frames are very interesting but cannot be shown here due to copyright reasons.

[2][Online]. Available: http://www.eink.com

[3][Online]. Available: http://www.hitachi.com

[4][Online]. Available: http://www.lgdisplay.com/

[5][Online]. Available: http://www.frontech.fujitsu.com/services/products/paper/flepia/

B. Related Work

Digital halftoning [7], [19], [20], [22], [27] refers to the physical process of converting a continuous tone image into a special image format, halftone image, which is composed of black and white dots, as shown in Fig. 1. We can find that the halftone image approximately keeps the visual characteristic of the continuous tone image at a proper viewing distance. In particular, video halftoning is a means of transferring a general video sequence into a format that can be displayed on devices with limited intensity resolutions and color palettes. In the literature, only a few video halftoning techniques have been discussed. In [9], a 3-D error diffusion algorithm that is used to mitigate the "flickering" flaw was proposed. Gotsman [8] applied an iterative image halftoning algorithm to deal with the problem of temporal flickering existing between video frames. This algorithm achieves visual results better than those obtained from classic halftoning algorithms at the expense of increasing computational load. In [1], Atkins *et al.* investigated the display of color image sequences using a model-based approach for multilevel error diffusion. Their algorithm achieved improvement in image quality over that yielded by frame-independent quantization when the frame rate is sufficiently high to support temporal flickering by the human visual system. In [10], a direct-binary-search algorithm was applied to 3-D error diffusion. In [26], Sun presented a motion-adaptive gain control-based 3-D error diffusion method to enhance the temporal consistency of the visual patterns by minimizing the flickering flaws.

From the existing methods, it is found that flickering flaws can be efficiently reduced from each frame. The basic ideal of flickering reduction is described as follows. Before a halftone video frame is produced, its original grayscale counterpart is compared with the previous one to decide the halftone values. This is because two neighboring halftone values along the temporal direction may create flickering. If the difference between two neighboring-frame pixels is small, then the halftone value of the pixel to be halftoned is decided from the previous halftone one; otherwise, the halftone value is determined from digital halftoning. This kind of video halftoning mechanism is regarded as neighboring-frame referencing-based video halftoning. Flickering reduction, however, usually accompanies visual quality degradation because some parts of a halftone frame will be made to be equal to its previous halftone one for flickering reduction. As a result, there is a tradeoff between the average temporal frequency of flickering and the quality of halftone video.

C. Our Contribution

In this paper, our purpose is to first propose temporal frequency of flickering-distortion (TFoFD) as a metric for performance evaluation of video halftoning and then propose a TFoFD optimized video halftoning method. In order to evaluate the performance of video halftoning, two metrics, "average temporal frequency of flickering (ATFoF)" and "weighted signal-to-noise ratio (WSNR)", are adopted, where ATFoF is used to measure visual quality along the temporal direction while WSNR is used to measure visual quality in the spatial domain. The tradeoff between ATFoF and WSNR is achieved via the proposed temporal frequency of flickering-distortion optimization mechanism.

Our major contributions are summarized here.

1) We uncover three major visual artifacts of conventional video halftoning methods.
2) We present a flickering sensitivity-based reference frame generation scheme for video halftoning with the capability of efficient flickering reduction while maintaining acceptable quality.
3) We propose a temporal frequency of flickering-distortion optimized video halftoning method with optimal flickering reduction threshold determination.
4) The proposed temporal frequency of flickering-distortion optimization mechanism can also be applied to the existing methods for performance improvement.

The remainder of this paper is organized as follows. In Section II, the flickering flaws in video halftoning and three major visual sensitivities, which are generated from flickering reduction, are described. In Section III, the proposed reference frame update per group of pictures (GOP)-based error diffusion and temporal frequency of flickering-distortion of video halftoning are described. Next, temporal frequency of flickering-distortion optimization of video halftoning will be proposed in Section IV. Extensive experimental results and comparisons are given Section V to demonstrate the effectiveness of our method. Finally, concluding remarks are summarized in Section VI.

II. PROBLEM STATEMENT

Here, we first discuss two kinds of flickering resulting from video halftoning in Section II-A. Then, we illustrate three major visual artifacts, which are generated from adopting conventional neighboring-frame referencing-based video halftoning methods for flickering reduction in Section II-B.

A. Flickering Flaws in Video Halftoning

A general video halftoning method consists of spatial error diffusion and temporal error diffusion, both of which create the flickering phenomena. Flickering is defined as the change of halftone values (either from black to white or from white to black) in the display of consecutive video frames that will be easily perceived by human eyes. For temporal error diffusion, this procedure will cause the pixels located at the same positions of neighboring video frames to have different halftone values due to the introduced diffused temporal errors, in particular, when the pixels have the same or similar gray values. For spatial error diffusion, the diffused spatial errors will affect the halftoning results of the subsequent grayscale pixels. If the area of grayscale pixels located at the same positions of neighboring video frames is affected by different diffused spatial errors, then the resultant halftone values may be different, leading to flickering flaws. This situation occurs with higher probabilities for pixels with grayscale values close to the quantization threshold (e.g., 128) of halftoning. Fig. 2 shows the effect of flickering flaws. Specifically, the white dots in Fig. 2(c) indicate the changes of halftone values, which will make the human eye feel uncomfortable when displaying Fig. 2(a) and (b) successively. In our study, the flickering flaws caused by either temporal error diffusion or spatial error diffusion are called unacceptable flickering, which has to be eliminated. Nevertheless,
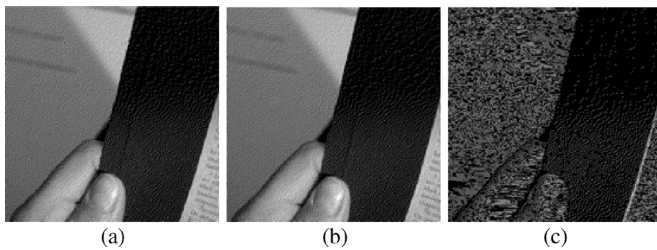
Fig. 2.   Flickering flaws. (a), (b) Neighboring halftone video frames. (c) Differences, illuminated with white dots (i.e., flickering flaws), between (a) and (b).

Fig. 3.   Real example of three artifacts generated from neighboring frame referencing-based video halftoning for flickering reduction: (left ellipse) residual shadow on hairs; (middle ellipse) subtle motion covered at the face; (right ellipse) dirty blocks on the coat/hands.
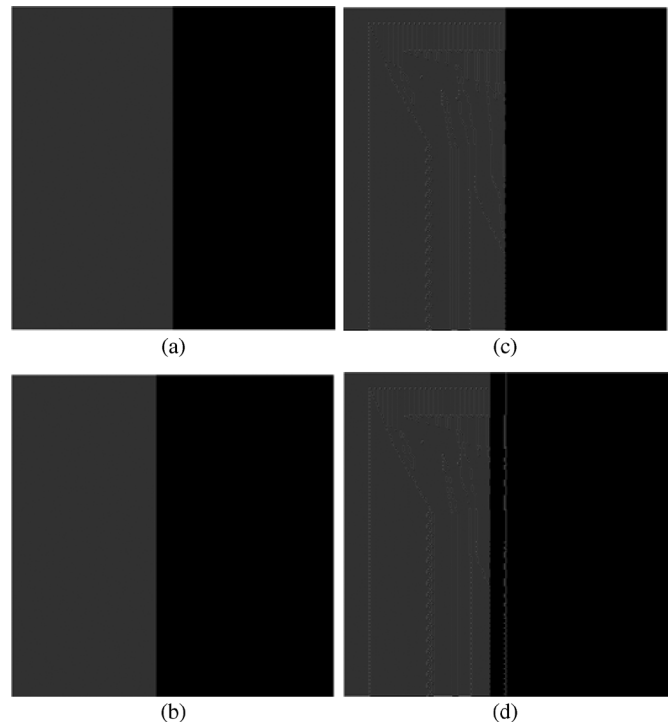

Fig. 4. Example of residual shadow: (a) grayscale frame, (b) grayscale frame next to (a), (c) halftone frame of (a) generated by error diffusion, and (d) halftone frame of (b) generated by neighboring-frame referencing-based video halftoning for flickering reduction.

not all flickering destroys the visual quality of a halftone video. Some is necessary for representing frame differences (motions) in digital halftone videos. We call this kind of flickering flow acceptable flickering, which contrasts with unacceptable flickering. Under this circumstance, the pixels located at the same spatial locations between two neighboring video frames can have different halftone values.

B. Three Visual Artifacts Generated From Flickering Reduction

Here, we introduce three visual phenomena, which appear after neighboring frame referencing-based video halftoning is applied for flickering reduction. As shown in Fig. 3,[6] the three visual phenomena illustrated with ellipses easily attract visual sensitivity.

1) Residual Shadow: Residual shadow is a visual phenomenon where the edge of a previous grayscale frame remains in a halftone frame after video halftoning. According to error diffusion halftoning, the white and black pixels are uniformly distributed on a halftone image to represent the illumination of corresponding gray pixels. Under this circumstance, an edge is

---

[6]This clip is available for viewing online at http://0rz.tw/su9TS http://www.iis.sinica.edu.tw/~cyhsu/index.files/Page695.htm.

blurred after it is halftoned by error diffusion, which means that the illumination of the edge's brighter side is diffused across the edge and the illumination of the other side is increased by redundant white halftone values. Consider a neighboring-frame referencing-based video halftoning method used for flickering reduction. The redundant white dots appear to form a shadow after flickering reduction even if the edge moves alone the inverse direction of error diffusion. An example of residual shadow is illustrated in Fig. 4. Fig. 4(a) and (b) shows two neighboring grayscale frames, where it can be observed that the edge moves from the right to the left. A halftone frame generated from Fig. 4(a) via error diffusion is shown in Fig. 4(c), where the illumination is diffused across the edge as we have discussed. To reduce flickering and computational complexity, the halftone values of the next halftone frame generated from Fig. 4(b) are decided from its previous halftone frame in Fig. 4(c) if the difference between two corresponding grayscale pixels is negligible. Fig. 4(d) shows the halftone frame corresponding to Fig. 4(b) when neighboring-frame referencing-based video halftoning is employed. It is found that the edge in Fig. 4(c) is kept as a shadow in Fig. 4(d). The shadow will hold on the video until the content is greatly changed. As a real example, the leftmost ellipse of Fig. 3 illustrates a residual shadow.

2) Subtle Motions: Subtle motions mean that a slight and slow motion in a video frame is covered if neighboring-frame referencing-based video halftoning is adopted. To reduce unacceptable flickering, traditional methods compare two pixels at the same position between two neighboring frames and use the resultant difference to decide the halftone value. The halftone
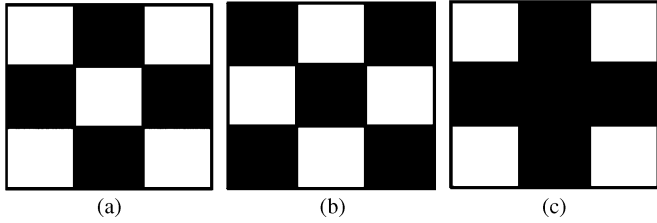
Fig. 5. Example of dirty block/spot: (a) error diffusion halftoned block, (b) halftone structure when the illumination of its grayscale version, when compared with the grayscale version of (a), is slightly decreased, and (c) the dirty block generated from conventional neighboring-frame referencing-based video halftoning methods.

Fig. 6. Block diagram of our video halftoning method.

value is decided to be the same as the previous one if the difference is smaller than the flickering reduction threshold. Thus, a halftone pixel continues to keep its state until the pixel difference is greater than the threshold. Based on the principle of neighboring-frame referencing-based video halftoning, the pixel difference that is caused due to motion between two neighboring frames may be ignored as being covered by the previous frame during video halftoning. For example, the video frames that present slow variations, such as the illumination changes of sunrise and sunset, possess slow motions. This kind of video clip will be displayed as a sequence composed of the same still halftone image, which is exactly the first frame produced after video halftoning, because the subsequent pixel differences are always smaller than the flickering reduction threshold. As a real example, the middle ellipse of Fig. 3 illustrates the portions whose slight and slow motions are covered during halftoning.

3) Dirty Block: We also find that dirty blocks/spots appear in halftone frames generated by neighboring frame referencing-based video halftoning. The spots are generated due to nonuniformly distributed halftone structures since the flickering between two frames is defined according to a flickering reduction threshold. Fig. 5 illustrates an example of dirty blocks. Fig. 5(a) and (b) shows two error diffusion frames with the illuminations of their corresponding grayscale frames differing slightly. Nevertheless, the flickering between Fig. 5(a) and (b) achieve the maximum. In order to reduce flickering, only one white pixel flipping on Fig. 5(a) is enough to display the change between two frames with similar grayscale values. Fig. 5(c) shows the result with the purpose of flickering reduction, but the frame looks like a spot and dirtier than Fig. 5(b) since the halftone structure is damaged and the illuminations are not uniformly distributed. As a real example, the rightmost ellipse of Fig. 3 illustrates a kind of dirty block.

III. REFERENCE FRAME UPDATE-BASED ERROR DIFFUSION AND TFoFD OF VIDEO HALFTONING

We propose in Section III-A a new video halftoning method using reference frame update per GOP-based error diffusion to leverage flickering reduction and visual quality in a halftone video. Here, we only present our algorithms based on halftoning grayscale videos. Our algorithms, however, can be extended to color video halftoning. Namely, when the color space is determined, our algorithms are applied on each channel of the color space. In Section III-B, implications from temporal
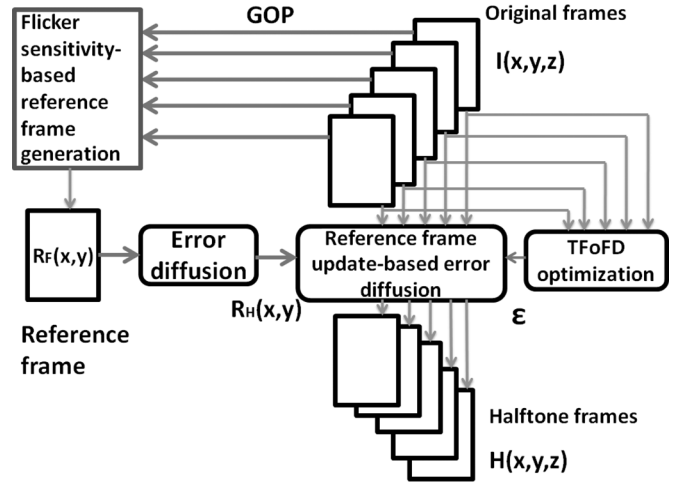
frequency of flickering-distortion of video halftoning will be obtained to figure out what an optimal flickering reduction threshold should be.

A. Reference Frame Update per GOP-Based Error Diffusion for Video Halftoning

In order to endow the proposed method with the ability to suppress visual artifacts, which we have described in Section II, the video frames are divided into many GOPs for video halftoning, similar to video coding. Each GOP consists of a reference frame (R_F frame) and a number of P frames, where P frames refer to the reference frame during the halftoning process. There may be two ways of achieving the segmentation of GOPs. The first one is similar to video shot change detection. The second one, which is adopted in this paper, refers to the change of average temporal frequency of flickering to determine the existence of new GOPs. In order to evaluate the performance of flickering reduction, the ATFoF between a pair of neighboring halftone video frames is defined as

$$\text{ATFoF}(z) = \sum_{x=1}^{X} \sum_{y=1}^{Y} |H(x,y,z) - H(x,y,z-1)| / (X \times Y) \tag{1}$$

which is used as a metric for indicating the degree of flickering flaw. In (1), $X \times Y$ denotes the size of a video frame and z denotes the frame index. If ATFoF ($0 \leq \text{ATFoF} \leq 1$) is larger than a threshold, then the frame z is regarded as the first frame of a new GOP.

Fig. 6 shows the block diagram of the proposed video halftoning method with GOPs being taken into consideration. Compared with the existing video halftoning methods, our method avoids adopting 3-D error diffusion. As shown in Fig. 6, to reduce unacceptable flickering, which we have described in Section II, a reference frame is generated after flickering sensitivity-based reference frame generation is conducted on a GOP. The reference frame $R_F(x,y)$ is then halftoned by spatial error diffusion [5] to generate the halftone reference frame $R_H(x,y)$. The subsequent halftone video frames in a GOP are not generated by means of spatial error diffusion. On the contrary, each of the subsequent video frames is compared
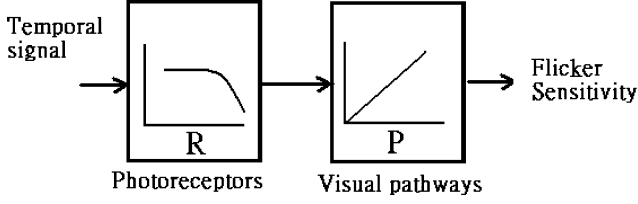
Fig. 7. Flicker sensitivity visual model simplified from [25].

with its reference frame during the subsequent halftoning process. For example, the zth grayscale frame $I(x, y, z)$ is compared with $R_F(x, y)$ for reference frame update-based error diffusion for video halftoning. If the gray-level difference between two pixels located at the same position is smaller than a flickering reduction threshold (to be described and optimized in Section IV), then the halftone value of the pixel in the zth frame is assigned[7] to be the same as that in $R_H(x, y)$; otherwise, the halftone value is determined via a quantization process in error diffusion. No matter which condition occurs, the quantization errors still need to be diffused. The above reference frame update-based error diffusion can be summarized as

$$
\begin{aligned}
&H(x, y, z) \\
&= \begin{cases} R_H(x, y), & \text{if } |I(x, y, z) - R_F(x, y)| \leq \epsilon \\ \text{halftoning}\,(I(x, y, z)), & \text{otherwise} \end{cases}
\end{aligned}
\tag{2}
$$

where function halftoning(.) can be any digital halftoning method and ϵ[8] is the flickering reduction threshold that will be optimally determined later in Section IV. In the proposed video halftoning method, video frames are halftoned sequentially. In each frame, the pixels are processed in the order, from left to right and top to bottom.

In the next two subsections, we shall describe how to generate a reference frame for video halftoning. First, a human flickering sensitivity model will be introduced. Second, based on the model, a flickering sensitivity-based clustering will be developed to generated reference frames.

1) Human Visual Model-Based Flickering Sensitivity: One important issue that should be emphasized here is how to systematically determine the flickering sensitivity threshold ϵ_{FST} used for reference frame generation. Since the observation (or existence) of flickering flaws depends on the sensitivity of human eyes to flickering, it is reasonable to make use of the models of human temporal sensitivity to derive ϵ_{FST}. There are a number of such models available in the literature [6], [25]. In particular, we find Rovamo *et al.*'s model [25] to be closest to the real visual model. Therefore, the threshold, ϵ_{FST}, is determined based on Rovamo *et al.*'s flickering sensitivity model, as shown in Fig. 7.

---

[7]In fact, this is similar in principle to block-based compression with "skip mode" for bit-rate saving. Therefore, our halftoning processing also has potential to benefit halftone video compression [14].

[8]A single threshold is considered here to simplify the computation complexity of finding the optimal threshold, to be described later.

In [25], the human visual system is modeled as a temporal signal processing system, which includes two major components: the modulation transfer function of the retina and the high-pass filtering in the postreceptoral retina. In Fig. 7, the frequency response $P(f)$ is defined as

$$
P(f) = af \tag{3}
$$

where f denotes a frequency variable and a is scaled to unity for simplicity. In addition, the impulse response of $R(f)$ is defined as

$$
r(t) = exp\left(-\left(\frac{ln\left(\frac{t}{\tau}\right)}{\sigma}\right)^2\right) \tag{4}
$$

where τ and σ are parameters that determine the peak position and width of the impulse response, respectively. Based on the two frequency responses $P(f)$ and $R(f)$, the frequency response $H(f)$ of the flickering sensitivity model is expressed as

$$
H(f) = R(f)P(f) \tag{5}
$$

where $R(f)$ is the frequency response of $r(t)$. Based on the Fourier transform of (5), and plugging it into (3), we can derive the impulse response of the flickering sensitivity model as

$$
\begin{aligned}
h(t) &= \int_{-\infty}^{\infty} H(f)e^{j\omega t} df \\
&= \int_{-\infty}^{\infty} R(f)P(f)e^{j\omega t} df \\
&= \int_{-\infty}^{\infty} R(f)af\frac{j2\pi f}{j2\pi f}e^{j\omega t} df \\
&= \frac{a}{j2\pi}\int_{-\infty}^{\infty} j2\pi f R(f)e^{j\omega t} df.
\end{aligned}
\tag{6}
$$

Then, (6) can be rewritten on the basis of Fourier transform, i.e., the differentiation in the time domain is replaced with that of multiplication by $j2\pi f$ in the frequency domain, and (4) as

$$
\begin{aligned}
h(t) &= \frac{a}{j2\pi}\int_{-\infty}^{\infty} j2\pi f R(f)e^{j\omega t} df \\
&= \frac{a}{j2\pi}\frac{\partial}{\partial t}r(t) \\
&= \frac{j2aln\left(\frac{t}{\tau}\right)}{t\sigma}exp\left(-\left(\frac{ln\left(\frac{t}{\tau}\right)}{\sigma}\right)^2\right).
\end{aligned}
\tag{7}
$$

With the flickering sensitivity model-based impulse responses [(7)] [25] and a video sequence $(I(x, y, z))$ available, the output $conv(I(x, y, z), h(z))$ of the flickering sensitivity model, denoted as the convolution between $I(x, y, z)$ and $h(z)$, represents the actual illumination of a video sequence sensed by human eyes along the temporal direction z. Fig. 8 illustrates
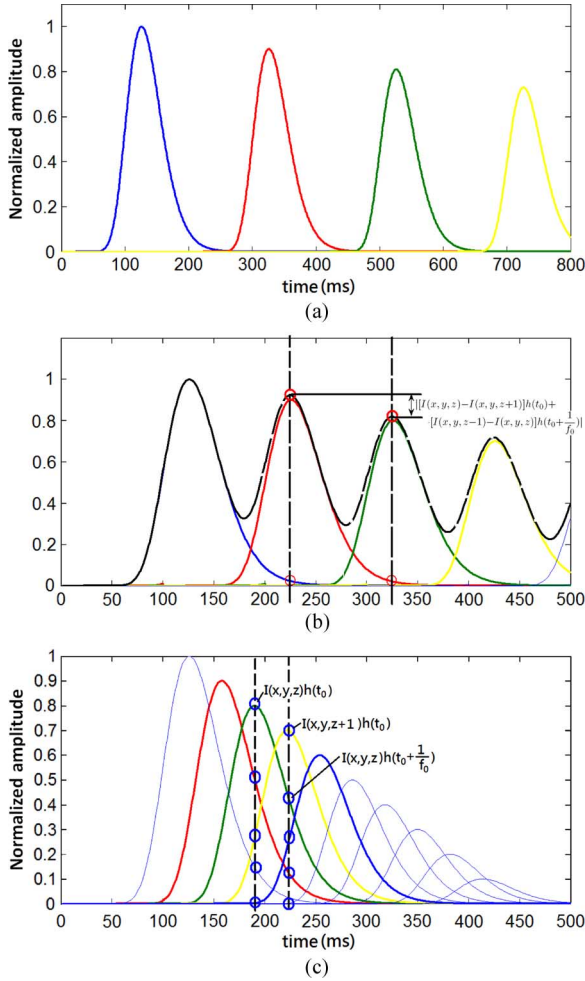
Fig. 8. Output of the flicker sensitivity model [25] that is the result of convolving the impulse response of the flicker sensitivity model with a video sequence at three different frame rates. The dashed curve in (b) denotes the sum of model outputs along the temporal direction. (a) 5 fps. (b) 10 fps. (c) 30 fps.

the model outputs (abbreviated from outputs of the flickering sensitivity model) at three different frame rates that generally can cover all cases in stating the relationship between model output and flickering sensitivity threshold. More specifically, the flickering sensitivity threshold ϵ_{FST} can be derived at location (x, y) from the difference between two aggregated model outputs, that is,

$$\left| \sum_n I(x, y, z + n) h\left(t_0 + \frac{n}{f_0}\right) - \sum_n I(x, y, z + (n+1)) h\left(t_0 + \frac{n}{f_0}\right) \right| \leq d \quad (8)$$

where d is a constant and represents the maximum flickering energy that can be ignored by human eyes, t_0 represents the maximum response of the flickering sensitivity model, and f_0 is the frame rate of the video sequence. In addition, on the left-hand side of (8), each term is a sum of n model outputs coming from n frames at the same time instance.

In the following, we describe how the flickering sensitivity threshold can be derived from the three general cases shown in Fig. 8. When the frame rate is rather low (e.g., as low as

5 fps), as an example shown in Fig. 8(a), the flickering sensitivity threshold ϵ_{FST} can be easily derived as

$$|I(x, y, z) - I(x, y, z + 1)| \leq \frac{d}{h(t_0)} = \epsilon_{FST} \quad (9)$$

where only one frame is contained in each model output.

While the frame rate of a video sequence is middle (e.g., 10 fps), as shown in Fig. 8(b), (8) can be rewritten as

$$\left| [I(x, y, z) - I(x, y, z + 1)] h(t_0) + [I(x, y, z - 1) - I(x, y, z)] h\left(t_0 + \frac{1}{f_0}\right) \right| \leq d. \quad (10)$$

Therefore, the flickering sensitivity threshold can be derived by directly expanding (10) as

$$\epsilon_{FST}^- = \frac{-d - [I(x, y, z - 1) - I(x, y, z)] h\left(t_0 + \frac{1}{f_0}\right)}{h(t_0)}$$
$$\leq I(x, y, z) - I(x, y, z + 1)$$
$$\leq \frac{d - [I(x, y, z - 1) - I(x, y, z)] h\left(t_0 + \frac{1}{f_0}\right)}{h(t_0)}$$
$$= \epsilon_{FST}^+. \quad (11)$$

It is clear that flickering is insensible if the difference between two neighboring pixels falls into the interval $[\epsilon_{FST}^-, \epsilon_{FST}^+]$.

When a higher frame rate (e.g., as high as 30 fps) is considered, as shown in Fig. 8(c), the difference between any two neighboring pixels will be nearly a constant based on the assumption that motion is linear over a short period of time, that is,

$$I(x, y, z + n) - I(x, y, z + (n+1)) \approx I(x, y, z) - I(x, y, z + 1). \quad (12)$$

Under this circumstance, (8) can be simplified as

$$|I(x, y, z) - I(x, y, z + 1)| \sum_n h\left(t_0 + \frac{n}{f_0}\right) \leq d \quad (13)$$

and then the flickering sensitivity threshold can be derived as

$$|I(x, y, z) - I(x, y, z + 1)| \leq \frac{d}{\sum_n h\left(t_0 + \frac{n}{f_0}\right)} = \epsilon_{FST}. \quad (14)$$

Although the flickering sensitivity threshold ϵ_{FST} can be used directly as a flickering reduction threshold in (2) for video halftoning with proper flickering elimination, the major visual artifacts introduced in Section II still exist. The main reason is that the use of ϵ_{FST} makes the distribution of halftone values nonuniform, as described in Section II-B3. For this reason, we avoid using the flickering sensitivity threshold as a flickering reduction threshold but adopt it instead for flickering sensitivity-based clustering to obtain the minimum average temporal frequency of flickering (mATFoF) in the next subsection.

2) Flickering Sensitivity-Based Reference Frame Generation: A simple and intuitive way to achieve the goal of generating reference frames is K-means clustering, that is, the flickering within the same cluster should be minimized while apparent flickering is allowed between clusters. Let a pixel sequence $P_{x,y}(z)$ of length Z be composed of pixel values at the fixed spatial position (x, y) of a video sequence along the
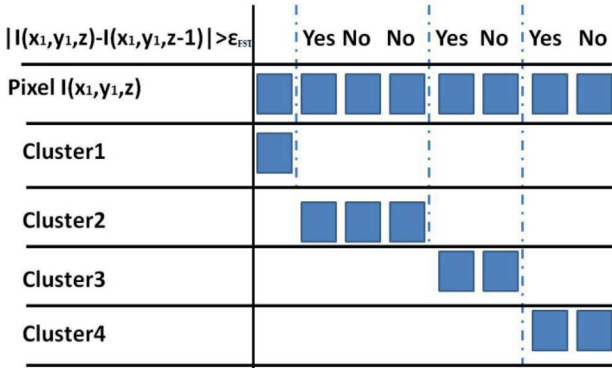
Fig. 9. Example of flickering sensitivity-based clustering.

temporal direction, where z denotes frame index, Z denotes the number of frames in a video, and $1 \leq z \leq Z$. K-means clustering aims to partition the n pixels into K clusters ($K < Z$), $C = \{C_1, C_2, \ldots, C_K\}$, so as to minimize the within-cluster sum of squares. In order to alleviate the flickering, the centroid of the largest cluster can be used to represent the pixel at position (x, y) of a generated reference frame. Nevertheless, it is found that the difference between the reference frame and each frame in a cluster is always larger than that between two neighboring frames.

In view of this, a flickering sensitivity-based clustering technique is proposed to restrain the increase of flickering as much as possible. Different from the K-means algorithm, flickering sensitivity-based clustering does not rely on iterative refinement. As shown in (11), the difference between two pixels located at the same position in temporal neighboring frames is calculated and compared with the sensitivity threshold ϵ_{FST} to determine the presence/absence of a flickering. Fig. 9 illustrates how flickering sensitivity-based clustering is performed. As shown in Fig. 9, the first pixel is directly assigned to the first cluster without needing to perform (11). Next, the pixel difference at position (x, y) between two neighboring frames for $1 < z \leq Z$ is subsequently calculated and compared with ϵ_{FST}; i.e., (11) is performed. If the difference is larger than ϵ_{FST}, a new cluster is generated; otherwise, a new pixel is added in the current cluster. This procedure repeats until all pixels in the pixel sequence $P_{x,y}(z)$ are assigned to clusters. Based on the proposed flickering sensitivity-based clustering, the centroid of the largest cluster (with the maximum number of pixels in it) for a fixed position (x, y) is calculated and used to represent the pixel value at (x, y) of the generated reference frame R_F.

It should be noted that the types of cluster at (x, y) along the temporal direction is decided by human visual model-based flickering sensitivity so as to guarantee that flickering does not appear between two temporal neighboring pixels in each cluster but only appear between two temporal neighboring clusters. Therefore, the minimum number of flickering at (x, y) is equal to the number of clusters. Furthermore, the minimum (mATFoF) per GOP, defined as the total number of clusters over the total number of pixels in a GOP, is also the one that can be achieved for video halftoning using flickering sensitivity-based clustering (mATFoF will be shown later).

We also observe that excessively restrained the flickering, however, will degrade halftone video quality in term of motion sharpness. In order to solve this problem, implications obtained from temporal frequency of flickering-distortion curves of video halftoning in Section III-B will reveal what an optimal flickering reduction threshold should be. In Section IV, we will use the mAFoF as a lower bound to determine the optimum flickering reduction threshold, which is used for video halftoning without incurring the visual artifacts discussed in Section II-B.

B. Implication From TFoFD of Video Halftoning

Let $I(x, y, z)$ be an original continuous tone video, and let $H(x, y, z)$ be the halftone version of $I(x, y, z)$. The goal of video halftoning here is to generate $H(x, y, x)$ from $I(x, y, z)$ by minimizing the visual difference between $I(x, y, z)$ and $H(x, y, z)$ while retaining acceptable average temporal frequency of flickering. Since a video sequence can be regarded as a 3-D signal, which can be perceived both spatially and temporally, the visual difference can be characterized with spatial errors and temporal errors simultaneously.

For spatial errors, the distortion between a pair of an original video frame and its corresponding halftone frame is calculated, i.e., the weighted signal-to-noise ratio (WSNR) [26] defined as

$$\mathrm{WSNR}(I, H) = 10 \log \frac{\sum_{x,y} [h(x, y) \otimes I(x, y, z)]^2}{\sum_{x,y} [h(x, y) \otimes (I(x, y, z) - H(x, y, z))]^2} \quad (15)$$

is adopted.

For temporal errors, the visual difference due to changes of video frames will raise the sensitivity of human eyes to temporal flickering. Therefore, we propose to use ATFoF to characterize the degree of temporal errors, which is defined as

$$\mathrm{ATFoF}(H) = \sum_{z=2}^{Z} \sum_{x=1}^{X} \sum_{y=1}^{Y} \frac{|H(x, y, z) - H(x, y, z-1)|}{X \cdot Y \cdot (Z-1)}. \quad (16)$$

In this paper, TFoFD optimization is proposed as a metric for evaluating video halftoning methods. The goal is to make sure that the perceptual quality of a halftone video is not destroyed by unconditionally reducing average temporal frequency of flickering.[9] It should be noted that the spatial-temporal WSNR [26] can be treated as a combination of two metrics based on two different filters for halftone video quality evaluation. For the low-pass filter case, the spatial temporal WSNR is a measurement of the filtered signal energy to filtered halftone distortion energy ratio. On the other hand, for the bandpass filter case, the spatial temporal WSNR measures the ratio of filtered signal energy to flicker distortion energy. We also used spatial temporal WSNR for measuring halftone video quality and observed that the spatial-temporal WSNR metric, depending on either low-pass filter or bandpass filter, cannot sufficiently represent the visual quality of a halftone video because the adopted filter

---

[9]Please note that the average temporal frequency of flickering will not be unconditionally increased if WSNR is minimized, as later shown in Fig. 10.
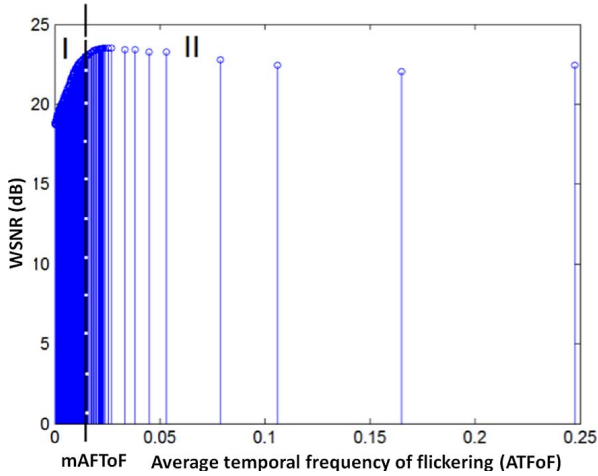
Fig. 10. Example of TFoFD curve obtained from observing a video clip. Similar phenomena can also be found in other video data.

can only filter parts of distortions out, that is, the objective spatial-temporal WSNR metric is not consistent with the subjective visual inspection. In addition, the computation of spatial-temporal filtering is too complex to be adopted in our proposed temporal frequency of flickering-distortion optimization strategy. In view of the above two concerns, it is better and reasonable to adopt our proposed TFoFD metric for performance evaluation and comparison.

Recall that the proposed reference frame update-based error diffusion is defined in (2). Apparently, the threshold ϵ controls the mutual behavior between average temporal frequency of flickering and distortion. Fig. 10 illustrates an example of TFoFD curve obtained from observing a GOP of video. Each pair of average temporal frequency of flickering and distortion corresponding to a threshold ϵ (ranging from 0 to 255) is represented as a point in Fig. 10. Similar phenomena can also be observed from other video sequences.

The meaning behind this kind of TFoFD curve can be interpreted from two parts, which are separated by the mATFoF, which has been described in the above subsection. The first part is located on the left-hand side of mATFoF, where the visual quality (in terms of WSNR) rapidly increases with slowly increased ATFoF. As shown in Fig. 10, most of the pairs of ATFoF and distortion are located within the first part. In the second part, the video qualities basically keep steady with increasing ATFoFs. As we have described in Section III-A2, the mATFoF can be obtained if flickering sensitivity clustering-based video halftoning is employed. As a result, it can be known that the best quality (in terms of WSNR) is located on the right-hand side of mATFoF.[10] We define the optimal TFoFD point as the point with the best tradeoff between WSNR and ATFoF.

The phenomena revealed in Fig. 10 are understandable. Recall that a larger flickering reduction threshold leads to a smaller average temporal frequency of flickering, as indicated in (2). Therefore, in the first part, a large threshold makes most of the flickering, including the acceptable flickering for motion compensation between two frames, be restrained. In particular, when the largest threshold (255) is used, the ATFoF becomes zero and

---

[10]Recall that, as we have described in Section III-A2, the quality of a halftone video corresponding to mATFoF is not the best.

the generated halftone frame is exactly the same as the reference halftone frame, leading to the distortion between the halftone frame and its corresponding original being the largest. When the threshold ϵ deviates from 255, the original frame (rapidly) becomes similar to the one obtained via halftoning and the distortion between them decreases accordingly since acceptable flickering are permitted. On the other hand, we also observe that the WSNR values in the second part do not significantly either increase or decrease with increasing average temporal frequency of flickering. In particular, the distortion first achieves the minimum and then gradually becomes static no matter what the average temporal frequency of flickering is.

From the above observations, we define the flickering reduction threshold as the one that can yield the highest WSNR value. In Section IV, we propose a TFoFD optimization method to find the optimal flickering reduction threshold.

IV. TFoFD OPTIMIZATION

Here, a TFoFD estimation method is proposed to find an optimum flickering reduction threshold (ϵ_{opt}), which can best achieve the tradeoff between visual quality and average temporal frequency of flickering. To estimate the TFoFD curve, we first define the probability modal of video halftoning in Section IV-A. Based on the proposed probability model, average temporal frequency of flickering and distortion, respectively, are estimated from an original video in Section IV-B and Section IV-C. The estimated average temporal frequency of flickering and distortion are combined to generate an estimated TFoFD curve for optimization in Section IV-D. It should be noted that the study of video halftoning from the viewpoint of TFoFD optimization has not been found in the literature.

A. Probability Model of Digital Halftoning

In order to analyze and estimate the distortion and average temporal frequency of flickering resulted from a halftone video, a halftone value is defined as a Bernoulli random variable, which takes either value 255 (denoted as a white dot) with probability p or value 0 (denoted as a black dot) with probability $q = 1 - p$. Since, the aim of image halftoning is to simply use only white and black dots to represent a gray-level picture, the average illumination of a halftone frame must be exactly equal to that of the original gray-level frame for reservation of perceptual quality in terms of average illumination. Therefore, the probability of a halftone value assigned to be a white dot is directly proportional to its corresponding grayscale pixel's illumination. With this understanding, the appearance probability of a halftone value can be modeled as

$$
\begin{aligned}
p(x,y,z) &= P\left(H(x,y,z) = 255\right)\\
&= 1 - P\left(H(x,y,z) = 0\right)\\
&= \frac{I(x,y,z)}{255}.
\end{aligned}
\tag{17}
$$

Thus, the distortion between the original frame $I(x,y,z)$ and halftone video frame $H(x,y,z)$ at position (x,y) of frame z can be estimated as

$$
\begin{aligned}
|I(x,y,z)-H(x,y,z)|^2 = {}&p(x,y,z)\cdot|I(x,y,z)-255|^2\\
&+(1-p(x,y,z))\cdot|I(x,y,z)|^2
\end{aligned}
\tag{18}
$$

where the first/second term on the right-hand side of (18) is the distortion for pixels halftoned to be white/black dots. The probability map $p(x, y, z)$ is also used to estimate the flickering's appearance in the next section.

B. ATFoF Estimation

For ATFoF estimation, we first define the appearance probability of flickering at the pixels located at (x, y) between two neighboring frames as

$$
\begin{aligned}
P\left(H(x, y, z) \neq H(x, y, z-1)\right) \\
= P\left(H(x, y, z) = 255\right) P\left(H(x, y, z-1) = 0\right) \\
+ P\left(H(x, y, z) = 0\right) P\left(H(x, y, z-1) = 255\right) \\
= p(x, y, z) \cdot (1 - p(x, y, z-1)) \\
+ p(x, y, z-1) \cdot (1 - p(x, y, z)).
\end{aligned}
\tag{19}
$$

According to (19), the estimated average temporal frequency of flickering as a function of the flickering reduction threshold ϵ is defined in (20), shown at the bottom of the page, where XY denotes the frame size and Z denotes the number of frames in a video. Equation (20) will be associated with the function of distortion estimation (described next) to find the estimated temporal frequency of flickering-distortion curve.

C. Distortion Estimation

The estimated distortion due to video halftoning comes from two cases. The first case includes the distortion generated from the pixel differences smaller than ϵ. Under this circumstance, the distortion resulted from halftoning at the position (x, y) of frame z can be derived as

$$
\begin{aligned}
D(x, y, z) = P\left(R_H(x, y) = 255\right) \cdot |I(x, y, z) - 255|^2 \\
+ P\left(R_H(x, y) = 0\right) \cdot |I(x, y, z)|^2.
\end{aligned}
\tag{21}
$$

Recall that the halftone value is either 255 or 0. According to (17), (21) can be rewritten as

$$
\begin{aligned}
D(x, y, z) = \frac{R(x, y)}{255} \cdot |I(x, y, z) - 255|^2 \\
+ \left(1 - \frac{R(x, y)}{255}\right) \cdot |I(x, y, z)|^2.
\end{aligned}
\tag{22}
$$

The second case consists of distortions generated from the pixel difference larger than ϵ. Under this circumstance, the distortion is independently generated from digital halftoning, as indicated in

$$
\begin{aligned}
D(x, y, z) = p(x, y, z) \cdot |I(x, y, z) - 255|^2 \\
+ (1 - p(x, y, z)) \cdot |I(x, y, z)|^2.
\end{aligned}
\tag{23}
$$

Thus, the overall distortion function depending on ϵ can be rewritten according to (22) and (23) as (24), shown at the bottom of the page.

Finally, incorporating (15) and (24), the WSNR with ϵ as the argument can be defined as

$$
\text{WSNR}(\epsilon) = 10 log \frac{\sum_{x,y} [h(x, y) \otimes I(x, y, z)]^2}{\sum_{x,y} [h(x, y) \otimes D(x, y, z, \epsilon)]^2}
\tag{25}
$$

where $\otimes$ denotes spatial convolution.

D. TFoFD Estimation for Optimum Flickering Reduction Threshold Selection

In (25), we have derived the relationship between distortion (in terms of WSNR) and ϵ. In this subsection, we first derive the relationship between average temporal frequency of flickering and ϵ, and then derive the relationship between ATFoF and distortion. Based on (20), the flickering reduction threshold-dependent ATFoF can be approximately rewritten as the percentage of pixel difference larger than ϵ multiplied by the expectation value of the average temporal frequency of flickering without adopting flickering reduction threshold, that is

$$
\begin{aligned}
\text{ATFoF}_{\text{esti}}(\epsilon) &\approx \frac{P(D_f > \epsilon)}{XY(Z-1)} \\
&\times \sum_{x,y,z} P\left(H(x, y, z) \neq H(x, y, z-1)\right) \\
&= P(D_f > \epsilon) E_{\text{ATFoF}}
\end{aligned}
\tag{26}
$$

where D_f is a random variable, which represents the absolute value of difference between two temporal neighboring pixels, and $E_{\text{ATFoF}} = (1/XY(Z-1)) \sum_{x,y,z} P(H(x, y, z) \neq H(x, y, z-1))$. It is well known that the probability distribution function of the pixel differences between two neighboring frames is a Laplacian distribution and that of absolute pixel

$$
\text{ATFoF}_{\text{esti}}(\epsilon) = \frac{1}{XY(Z-1)} \sum_{(x,y,z) \in \{(x,y,z) \mid |I(x,y,z) - I(x,y,z-1)| > \epsilon\}} P\left(H(x, y, z) \neq H(x, y, z-1)\right)
\tag{20}
$$

$$
D(x, y, z, \epsilon) = \begin{cases} \frac{R(x,y)}{255} \cdot |I(x, y, z) - 255|^2 + \left(1 - \frac{R(x,y)}{255}\right) \cdot |I(x, y, z)|^2, & \text{if } |I(x, y, z) - R(x, y)| \leq \epsilon \\ p(x, y, z) \cdot |I(x, y, z) - 255|^2 + (1 - p(x, y, z)) \cdot |I(x, y, z)|^2, & \text{otherwise} \end{cases}
\tag{24}
$$

differences is an exponential distribution. Therefore, (26) can be rewritten as

$$
\begin{aligned}
\text{ATFoF}_{\text{esti}}(\epsilon) &\approx P(D_f > \epsilon) \cdot E_{\text{ATFoF}} \\
&= (1 - P(D_f \le \epsilon)) \cdot E_{\text{ATFoF}} \\
&= \left(1 - (1 - e^{-\lambda\epsilon})\right) \cdot E_{\text{ATFoF}} \\
&= E_{\text{ATFoF}} \cdot e^{-\lambda\epsilon}
\end{aligned}
\tag{27}
$$

where λ can be calculated from the reciprocal of the mean of pixel values in the frame. From (27), we can derive ϵ as

$$
\epsilon = -\frac{ln\left(\frac{\text{ATFoF}_{\text{esti}}(\epsilon)}{E_{\text{ATFoF}}}\right)}{\lambda}.
\tag{28}
$$

Then, the relationship between average temporal frequency of flickering and distortion can be derived by substituting ϵ in (28) into (25), that is

$$
\begin{aligned}
&\text{WSNR}(\text{ATFoF}_{\text{esti}}) \\
&= 10log\frac{\sum_{x,y}\left[h(x,y) \otimes I(x,y,z)\right]^2}{\sum_{x,y}\left[h(x,y) \otimes D\left(x,y,z,-\frac{ln\left(\frac{\text{ATFoF}_{\text{esti}}(\epsilon)}{E_{\text{ATFoF}}}\right)}{\lambda}\right)\right]^2}.
\end{aligned}
\tag{29}
$$

Now, the TFoFD optimization can be used to find the threshold, which can lead to distortion minimization with acceptable ATFoF. As we have described in Section III-B, the best halftone video quality is located at the second part of the estimated TFoFD curve. We have observed from several results that only few points (less than 1/8 of total points) are located in the second part of the curve in Fig. 10. To find the optimum threshold ϵ_{opt}, we calculate each point in the second part of the temporal frequency of flickering-distortion curve and decide the ϵ_{opt} as the threshold that generates the best halftone video quality.

V. EXPERIMENTAL RESULTS

Extensive experiments were conducted to evaluate the effectiveness of the proposed reference frame update-based error diffusion video halftoning method for use in e-paper. In Section V-A, we evaluate the accuracy of the proposed methods in estimating the average temporal frequency of flickering, distortion, and temporal frequency of flickering-distortion under various flickering reduction thresholds. Next, the estimated results are employed to optimize the flickering reduction threshold in Section V-B. Finally, in Section V-C, we conduct performance comparisons among the proposed video halftoning method and some existing video halftoning methods, including the Floyd-Steinberg error diffusion method [5], ordered dither, 3-D error diffusion [9], and Sun's method [26].

In the experiments, a number of 24-b color video sequences were adopted. Among them, the results obtained from the "Vassar" sequence, the "Secretary" clip, which was excerpted from the movie "The Devil Wears Prada," and the "ballroom" video were reported here. These video sequences, respectively, contain small, moderate, and large motion contents. The frame size ranges from 480×640 to 436×1024.
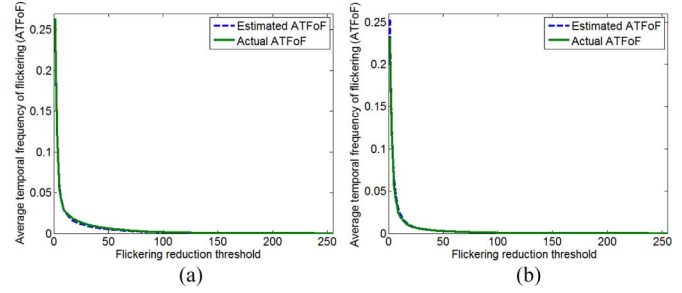

Fig. 11. Estimation of ATFoF via (20) for the video "Secretary." (a) Proposed reference frame update-based error diffusion. (b) modified 3-D error diffusion.
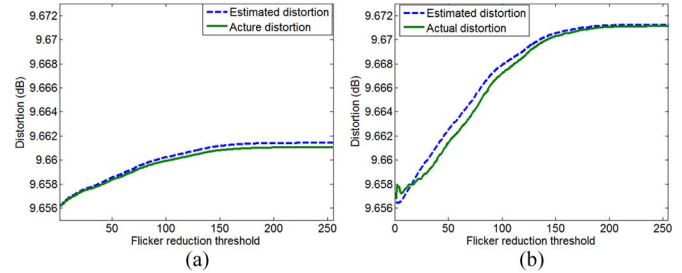

Fig. 12. Estimation of distortion via (24) for the video "Secretary." (a) Proposed reference frame update-based error diffusion. (b) Modified 3-D error diffusion.

A. Accuracy of the Estimated ATFoF, Distortion, and TFoFD

In order to verify whether the proposed estimations, including ATFoF, distortion, and TFoFD, are adaptive to and can be associated with the existing video halftoning methods. We evaluate the accuracy of the estimated temporal frequency of flickering [via (20)], estimated distortion [via (24)], and estimated temporal frequency of flickering-distortion [via (29)] under various flickering reduction thresholds. The evaluations were conducted by comparing the estimated results with the actual results produced from halftoning real video sequences. In our experiments, the aforementioned evaluation tasks (Section IV) were conducted on the proposed video halftoning method (Section III-A) and modified 3-D error diffusion.

It should be noted that the original 3-D error diffusion approach [9] adopts neighboring-frame referencing-based video halftoning strategy and only uses a fixed threshold to achieve video halftoning. In the experiment considered here, our proposed estimations (Section IV) are combined with 3-D error diffusion so that the optimal flickering reduction threshold can be determined.

The estimation and comparison results are shown in Figs. 11–13 for ATFoF, distortion, and TFoFD, respectively. We can observe from these results that the proposed TFoFD estimation mechanism exhibits rather accurate estimations in both two video halftoning methods. For other video sequences, similar results can also be obtained. As we have described, the goal of these experiments is to verify the accuracy and suitability of the proposed estimations. The results here indeed show that our proposed estimations are accurate and adaptive to the existing halftoning methods.

It should be noted that, although the proposed estimation mechanisms indeed improve 3-D error diffusion in reducing
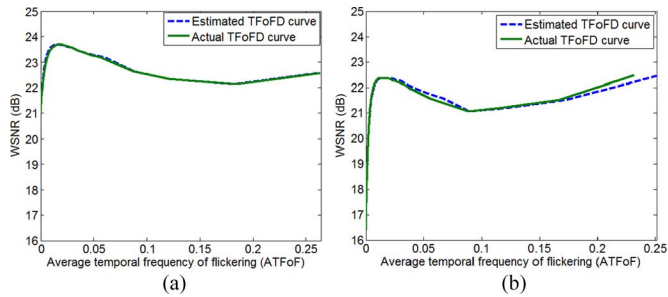
Fig. 13.   Estimation of TFoFD via (29) for the video "Secretary." (a) Proposed reference frame update-based error diffusion. (b) Modified 3-D error diffusion.
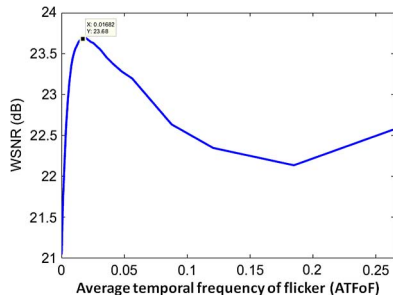

Fig. 14.   TFoFD optimization for video "Secretary." The peak shows the temporal frequency of flickering, 0.01682 and WSNR, 23.68 dB.
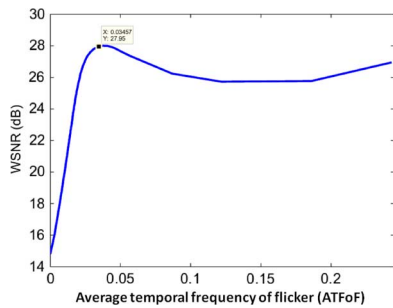

Fig. 15.   TFoFD optimization for video "Ballroom." The peak shows the temporal frequency of flickering, 0.03457, and WSNR, 27.95 dB.

flickering, as shown in Figs. 11–13, the three visual artifacts still remain due to the use of neighboring frame referencing-based halftoning. Also note that performance comparisons between the proposed video halftoning method and the original 3-D error diffusion will be described later in Section V-C.

B. Accuracy of TFoFD Optimization

To verify whether or not the proposed TFoFD optimization scheme can achieve the best tradeoff between the average temporal frequency of flickering and distortion based on an optimally determined flickering reduction threshold, the estimated optimal flickering reduction threshold (described in Section IV-D) was employed in the proposed video halftoning scheme to generate halftone videos. As with the optimal flickering reduction threshold, the resultant pair of ATFoF and distortion (in terms of WSNR) is labeled in the TFoFD curve, as shown in Figs. 14–16 for different videos. In this experiment, three video clips including small, moderate, and large motion contents were adopted. We can find from these results that the obtained pairs of average temporal frequency of flickering and
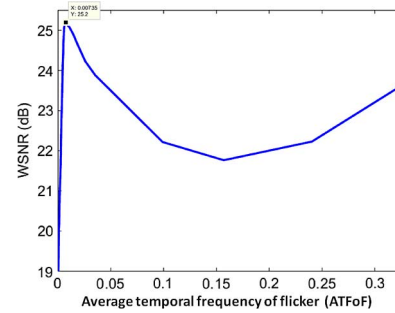

Fig. 16.   TFoFD optimization for video "Vassar." The peak shows the temporal frequency of flickering, 0.00735, and WSNR, 25.20 dB.

WSNR are located very near the peaks of the TFoFD curves, which implies that the proposed method is fairly accurate to determine the optimal flickering reduction threshold for video halftoning.

C. Performance Comparisons

In this experiment, different video sequences were used for comparisons between the existing video halftoning methods and our video halftoning method. The average temporal frequency of flickering and distortion of a halftone video were used as the metrics for performance evaluation. The four halftoning techniques used for comparison include the error diffusion [5], the ordered dither method, the 3-D error diffusion method [9], and Sun's video halftoning method [26]. Error diffusion and order dither were implemented based on Floyd-Steinberg error diffusion kernel and Bayer dither screen, respectively. The 3-D error diffusion approach and Sun's method were implemented according to the parameters in the original literatures. In addition, our proposed reference frame update per GOP-based method for flickering reduction and TFoFD optimization were applied to original order dither and 3-D error diffusion to study the introduced improvements. The average temporal frequency of flickering and WSNR obtained for each video sequence with respect to each method are shown in Table I,[11] from which we have the following observations.

First, the video qualities obtained using error diffusion-based video halftoning methods are overwhelmingly better than those obtained using order dither since the Bayer screen cannot uniformly disperse halftone pixels as efficiently as error diffusion. It can also be observed that the WSNR values of order dither and TFoFD optimized order dither are not as good as the others. Second, the optimized ordered dither and optimized 3-D error diffusion techniques obtain lower average temporal frequency of flickering and higher WSNR values for almost all video sequences used in the experiment. These results reveal the fact that the proposed optimal temporal frequency of flickering-distortion mechanism is able to find the flickering reduction threshold to optimally achieve the tradeoff between average temporal frequency of flickering and distortion. Third, Floyd–Steinberg error diffusion [5] achieves higher WSNR

[11]To access these clips, please go to http://0rz.tw/su9TS or directly key in http://www.iis.sinica.edu.tw/~cyhsu/index.files/Page695.htm for visual inspection of the halftone video clips obtained from the video halftoning methods used for comparisons.

TABLE I
COMPARISON OF VIDEO HALFTONING PERFORMANCE

Methods	Metrics	Vassar	Secretary	Ballroom	Trevor	Football
Ordered dither	WSNR(dB)	20.12	19.27	20.99	19.83	20.32
	ATFoF	0.012	0.017	0.027	0.013	0.027
Optimized ordered dither	WSNR(dB)	20.31	19.53	20.89	20.11	20.17
	ATFoF	0.003	0.016	0.020	0.009	0.021
3D error diffusion [9]	WSNR(dB)	23.69	21.33	25.67	22.91	26.13
	ATFoF	0.009	0.032	0.042	0.022	0.040
Optimized 3D error diffusion	WSNR(dB)	24.32	23.11	26.72	23.87	26.88
	ATFoF	0.003	0.026	0.033	0.016	0.034
Error diffusion [5]	WSNR(dB)	25.32	22.63	27.03	24.71	27.42
	ATFoF	0.152	0.263	0.242	0.239	0.239
Sun's method [26]	WSNR(dB)	23.61	22.32	23.93	23.12	23.81
	ATFoF	0.017	0.024	0.033	0.021	0.036
Proposed method	WSNR(dB)	25.31	23.73	27.95	24.63	26.38
	ATFoF	0.006	0.017	0.034	0.011	0.037

values for most cases but the average temporal frequency of flickering are too high to be acceptable. Fourth, we find that the two neighboring frame referencing-base video halftoning methods [9], [26] still exhibit visual artifacts that have been discussed in Section II-B, even if a flickering reduction strategy is applied. In sum, the experimental results indicate that our video halftoning method obtains acceptable temporal frequency of flickering while in most cases retaining WSNR values higher than the methods used for comparisons.

VI. CONCLUSION AND FUTURE WORK

In view of the fact that the video halftoning technology will play an important role in supporting the use of the emerging display device, e-paper, the issues of video halftoning are studied in this paper. The major contributions of this paper are summarized as follows: 1) we uncover three major visual artifacts of conventional video halftoning methods; 2) a flickering sensitivity-based reference frame generation scheme for video halftoning is proposed with capability of efficient flickering reduction while maintaining acceptable quality; 3) TFoFD optimization of video halftoning that is ignored in the literature is presented in this paper for optimal flickering reduction threshold determination; and 4) our TFoFD optimization strategy is verified to improve the performance of the existing video halftoning methods.

In the future work, a more advanced visual model for halftoning is required to further improve our method. First, it may be important to adopt the visual model defined in [21], which takes eye movement into account, which can be useful for flickering reduction. Second, a visual model can be introduced to design a probability model of digital halftoning, as we have described in Section IV We plan to study these issues that have not been covered in video halftoning.

REFERENCES

[1] C. B. Atkins, T. J. Flohr, D. P. Hilgenberg, C. A. Bouman, and J. P. Allebach, "Model-based color image sequence quantization," in *Proc. IS&T/SPIE Conf. Human Vis., Visual Process., Digit. Display V*, 1994, vol. 2179, pp. 310–317.
[2] P. J. Barten, "Physical model for the contrast sensitivity of the human eye," in *Proc. IS&T/SPIE Int. Symp. Electron. Imaging Sci. Technol.*, 1992, vol. 1666, pp. 57–74.
[3] B. E. Bayer, "An optimum method for two-level rendition of continuoustone pictures," in *Proc. IEEE Int. Conf. Commun.*, 1973, vol. 26, pp. 11–15.
[4] G. P. Crawford, "A bright new page in portable displays," *IEEE Spectrum*, vol. 37, pp. 40–46, 2000.
[5] R. Floyd and L. Steinberg, "An adaptive algorithm for spatial gray scale," in *Proc. Soc. Inf. Display*, 1976, vol. 17, no. 2, pp. 75–77.
[6] R. E. Fredericksen and R. F. Hess, "Estimating multiple temporal mechanism in human vision," *Vis. Res.*, vol. 38, pp. 1023–1040, 1998.
[7] J. Gille, J. Luszcz, and J. O. Larimer, "Improving the appearance of flat-panel displays using multilevel color error diffusion," in *Proc. SPIE/IST Conf. Human Vis. Electron. Imaging*, 1997, vol. 3016, pp. 154–167.
[8] C. Gotsman, "Halftoning of image sequence," *Vis. Comput.*, vol. 9, no. 5, pp. 255–266, 1993.
[9] H. Hild and M. Pins, "A 3-D error diffusion dither algorithm for halftone animation on bitmap screens," in *State-of-the-Art in Computer Animation-Proceedings of Computer Animation*. Berlin, Germany: Springer-Verlag, 1989, pp. 181–190.
[10] D. P. Hilgenberg, T. J. Flohr, C. B. Atkins, J. P. Allebach, and C. A. Bouman, "Least-squares model-based video halftoning," in *Proc. SPIE/IST Conf. Human Vis., Visual Process., Digit. Display V*, 1994, vol. 2179, pp. 7–10.
[11] C. Y. Hsu and C. S. Lu, "Joint image halftoning and watermarking in high-resolution digital form," in *Proc. IEEE Int. Conf. Multimedia Expo*, The Netherlands, 2005, pp. 534–537.
[12] C. Y. Hsu, C. S. Lu, and S. C. Pei, "Joint screening halftoning and visual cryptography for image protection," in *Proc. Int. Workshop Digital Watermarking*, Korea, 2006, vol. 4283, LNCS, pp. 212–225.
[13] C. Y. Hsu, C. S. Lu, and S. C. Pei, "Video halftoning preserving temporal consistency," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Beijing, China, 2007, pp. 1938–1941.
[14] C. Y. Hsu, C. S. Lu, and S. C. Pei, "Compression of halftone video for electronic paper," in *Proc. IEEE Int. Conf. Image Process.*, San Diego, CA, 2008, pp. 1600–1603.
[15] C. Y. Hsu, C. S. Lu, and S. C. Pei, "Scalable multi-layer halftone video display for electronic paper," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. 2008, pp. 1445–1448.
[16] *Standardization of Group 3 Facsimile Apparatus for Document Transmission*, ISO CCITT Recommend. T.4, 1980.
[17] *Coded Representation of Picture and Audio Information-Progressive Bi-Level Image Compression*, ISO/IEC JTC1 CD 11544, 1993.
[18] *JBIG2 Final Draft International Standard*, ISO/IEC JTC1/SC29/ WG1N1545, Dec. 1999.
[19] H. R. Kang, *Digital Color Halftoning*. New York: SPIE/IEEE Press, 1999.
[20] D. E. Knuth, "Digital halftones by dot diffusion," *ACM Trans. Graph.*, vol. 6, no. 4, pp. 245–273, 1987.
[21] J. Laird, M. Rosen, J. Pelz, E. Montag, and S. Daly, "Spatio-velocity CSF as a function of retinal velocity using unstabilized stimuli," in *Proc. SPIE/IST Conf. Human Vis. Electron. Imaging II*, 2006, vol. 6057, pp. 605705:1–605705:11.
[22] D. L. Lau and G. R. Arce, *Modern Digital Halftoning*. New York: Marcel Dekker, 2001.
[23] C. S. Lu and C. Y. Hsu, "Geometric distortion-resilient image hashing scheme and its applications on copy detection and authentication," *ACM Multimedia Syst. J.*, vol. 11, no. 2, pp. 159–173, Dec. 2005.
[24] Mitsubishi Electric Research Laboratories, MERL Multi-View Video Sequences [Online]. Available: ftp://ftp.merl.com/pub/avetro/ mvctestseq
[25] J. Rovamo, K. Donner, R. Nasanen, and A. Raninen, "Flicker sensitivity as a function of target area with and without temporal noise," *Vis. Res.*, vol. 40, pp. 3841–3851, 2000.

[26] Z. Sun, "Video halftoning," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 678–686, Mar. 2006.

[27] P. W. Wong and N. D. Memon, "Image processing for halftones," *IEEE Signal Process. Mag.*, vol. 20, pp. 59–70, 2003.

[28] J. Xin, A. Vetro, S. I. Sekiguchi, and K. Sugimoto, "Motion and mode mapping for MPEG-2 to H.264/AVC transcoding," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2006, pp. 313–316.

[29] [Online]. Available: http://www.sipix.com.tw

Chao-Yung Hsu is currently working toward the Ph.D. degree at the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan.

He has been a research assistant with the Institute of Information Science, Academia Sinica, Taipei, Taiwan, since 2003. His research interests include multimedia signal processing, data hiding, and digital halftoning.

Chun-Shien Lu (M'99) received the Ph.D. degree in electrical engineering from National Cheng-Kung University, Tainan, Taiwan, in 1998.

From October 1998 to July 2002, he was with the Institute of Information Science, Academia Sinica, Taipei, Taiwan, as a Postdoctoral Fellow for his military service. From August 2002 to July 2006, he was an Assistant Research Fellow with the same institute and, since July 2006, he has been an Associate Research Fellow. He was a guest co-editor of a special issue of *EURASIP Journal on Applied Signal Processing* on Visual Sensor Networks in 2005. He holds two U.S. patents, three R.O.C. patents, and one Canadian patent in digital watermarking. His current research interests mainly focus on various topics (including signal processing and security) of multimedia, sensor network security, and compressive sensing.

Dr. Lu is a member of the Association for Computing Machinery. He was the organizer of a special session on Multimedia Security in the 2nd and 3rd IEEE Pacific-Rim Conference on Multimedia, respectively (2001–2002). He co-organized two special sessions (in the area of media identification and DRM) in the 5th IEEE International Conference on Multimedia and Expo (ICME), 2004. He was the recipient of the Ta-You Wu Memorial Award, National Science Council, in 2007 and was corecipient of a National Invention and Creation Award in 2004. Since July 2007, he has served as a member of the Multimedia Systems and Applications Technical Committee of the IEEE Circuits and Systems Society. He is currently an associate editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.

Soo-Chang Pei (SM'89–F'00) was born in Soo-Auo, Taiwan, in 1949. He received the B.S.E.E. degree from the National Taiwan University, Taipei, Taiwan, in 1970, and the M.S.E.E. and Ph.D. degrees from the University of California, Santa Barbara, in 1972 and 1975, respectively.

Since 1984, he has been a Professor with the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan. He was an Engineering Officer with the Chinese Navy Shipyard from 1970 to 1971. From 1971 to 1975, he was a Research Assistant with the University of California, Santa Barbara. He was a Professor and the Chairman of the Electrical Engineering Department, Tatung Institute of Technology, Taipei, from 1981 to 1983, and was with the National Taiwan University as the Chairman of the Electrical Engineering Department from 1995 to 1998, and the Dean of the College of Electrical Engineering and Computer Science from 2003 to 2009, respectively. His research interests include digital signal processing, image processing, optical information processing, and laser holography.

Dr. Pei is a member of Eta Kappa Nu and the Optical Society of America. He was the President of the Chinese Image Processing and Pattern Recognition Society in Taiwan from 1996 to 1998. He became an IEEE Fellow in 2000 for his contributions to the development of digital eigenfilter design, color image coding, and signal compression and to the electrical engineering education in Taiwan. He was a recipient of a National Sun Yet-Sen Academic Achievement Award in Engineering in 1984, the Distinguished Research Award from the National Science Council from 1990 to 1998, an Outstanding Electrical Engineering Professor Award from the Chinese Institute of Electrical Engineering in 1998, the Academic Achievement Award in Engineering from the Ministry of Education in 1998, the Pan Wen-Yuan Distinguished Research Award in 2002, and the National Chair Professor Awards from the Ministry of Education in 2002 and 2008, respectively.