# A METHOD TO GENERATE HALFTONE VIDEO

*Zhaohui Sun*

Research & Development Laboratories
Eastman Kodak Company, Rochester, NY 14650-1816

## ABSTRACT

This paper studies video halftoning that renders digital video sequence onto display devices which have limited intensity resolutions and color palettes, by trading the spatiotemporal resolution for enhanced intensity/color resolution. Such a trade is needed when a continuous tone video is not necessary or not practical for video display, transmission, and storage. In particular, the quantization error of a pixel is diffused to its spatiotemporal neighbors by separable 1-D temporal and 2-D spatial error diffusions. Motion-adaptive gain control is employed to enhance the temporal consistency of the visual patterns by minimizing the flickering artifacts. Experimental results of halftone and colortone videos are demonstrated and evaluated with various halftoning techniques.

## 1. INTRODUCTION

Video halftoning is a task that renders video sequence onto display devices that have limited intensity resolutions and color palettes. Halftone video provides an alternative for video representation, rendering, storage, and transmission, when continuous tone video is not necessary or not practical, and it can be applied to display, data reduction, and error resilient communication. It can be used to render continuous tone video on display devices when there is a mismatch between the image/video representation and the display capability because of the constraints of cost and system complexity, such as small electronic gadgets, large screen display, and flexible display. With a shorter bit depth, the size of a halftone or colortone video is much smaller than its counterpart with continuous tone, and it can be further reduced after exploring the temporal consistency. As stochastic noise patterns are used to conceal the quantization errors in the spatiotemporal domain, any random perturbation on the halftone video, such as channel noise, is less pronounced in terms of image quality degradation. Therefore, it is particularly suitable for wireless communication.

Related prior art includes digital image halftoning and the extension to handling image sequences. Image halftoning reduces the intensity/color resolution of an image for the best possible reproduction and has wide applications in printing and display industries. Studies have also been carried out to apply the halftoning technology to image sequences [1, 2, 3]. A 3-D error diffusion algorithm is proposed in [1], including a constant gain control scheme to minimize temporal flickering. In [2], an iterative image halftoning algorithm is applied to image sequence where the halftone map on the previous frame is used as the starting point for iterative refinement on the current frame. In [3], spatiotemporal error diffusion filters are designed for the luminance and chrominance channels at different frame rates.

Motivated by the widespread use of image halftoning in the printing industry, we extend the concept to video halftoning in the display applications by reducing the video tone scale with the minimum visual degradation. The major contributions of this paper include a scheme of video tone scale reduction by separable temporal and spatial error diffusion and a method of temporal flicker reduction by the use of motion information.

## 2. SEPARABLE ERROR DIFFUSION

A digital video sequence $\mathbf{V} = \{I(i,j,k),\ i = 1 \ldots M,\ j = 1 \ldots N,\ k = 1 \ldots K\}$ is a temporally varying 2-D spatial signal $I$ on frame $k$, sampled and quantized at spatial location $(i,j)$ with $b$ bits per pixel. Video halftoning is a task to transform a full resolution video $\mathbf{V}$ with a continuous tone scale (e.g., $b = 8$) to a halftone video $\mathbf{V}_d$ with a lower bit depth $b_d < b$ (e.g., $b_d = 1$), such that the perceived visual difference is made as small as possible.

It has been shown in [4] that the 3-D spatiotemporal impulse response of the human visual system (HVS), although very complicated in nature, is separable in temporal and spatial dimensions in approximation. Therefore, the 3D spatiotemporal error diffusion can be carried out by a temporal error diffusion followed by a spatial error diffusion, which greatly simplifies the system complexity. The basic idea is to diffuse part of the quantization error at a pixel to its causal temporal neighbor along the motion trajectory and the rest to its casual spatial neighbors to minimize the spatial visual distortion. The exact amount is controlled by a temporal diffusion map.

The separable error diffusion scheme is presented in Fig. 1.

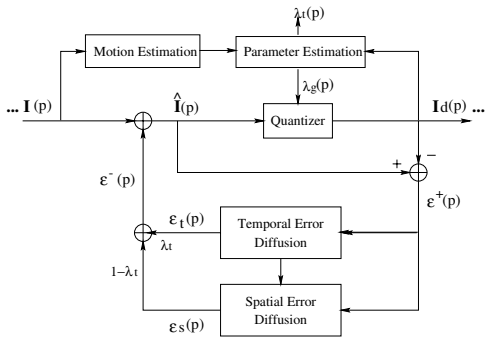


**Fig. 1**. Separable temporal and spatial error diffusion.

The video frames are processed sequentially. The pixels inside the frame are scanned in a serpentine order, from left to right on even lines and from right to left on odd lines. At a pixel location $p = (i, j, k)$, the image intensity $I(i, j, k)$ and the quantization errors diffused from its spatiotemporal neighbors $\varepsilon^-(i, j, k)$ are quantized to $I_d(i, j, k)$, by a comparison of $\hat{I}(i, j, k) = I(i, j, k) + \varepsilon^-(i, j, k)$ with the threshold $T(i, j, k)$ (e.g., $T = I_m$),

$$I_d(i,j,k) = \begin{cases} 0 & \text{if } I(i,j,k) + \varepsilon^-(i,j,k) < T(i,j,k) \\ 1 & \text{if } I(i,j,k) + \varepsilon^-(i,j,k) \geq T(i,j,k). \end{cases} \quad (1)$$

Pixel $p = (i, j, k)$ on halftone video $\mathbf{V}_d$ is denoted as a black dot if the adjusted intensity value is less than the threshold, or a white dot otherwise, yielding a quantization error

$$\varepsilon^+(i,j,k) = \hat{I}(i,j,k) - I_d(i,j,k). \quad (2)$$

To improve the visual quality, the quantization error $\varepsilon^+(i, j, k)$ is diffused to the spatiotemporal neighbors. As shown in Fig. 2(a), part of the error is diffused along the motion trajectory to the temporal neighbor as $\varepsilon_t$ and the rest to the intraframe neighbors as $\varepsilon_s$ in the spatial domain. The diffused error $\varepsilon^-$ in (1) are collected from its spatiotemporal neighbors as shown in Fig. 2(b),

$$\varepsilon^-(i,j,k) = \lambda_t(i,j,k) \cdot \varepsilon_t(i + d_x(i,j), j + d_y(i,j), k-1) + (1 - \lambda_t(i,j,k)) \cdot \sum_{s^i \in \mathcal{S}} \alpha_i(I(i,j,k)) \cdot \varepsilon_s(i + s_x^i, j + s_y^j, k). \quad (3)$$

Part of $\varepsilon^-(i, j, k)$ is contributed by $\varepsilon_t$ from the temporal neighbor on the previous frame with a weight of $\lambda_t(i, j, k)$, and the rest from $\varepsilon_s$ from the spatial neighbors on the current frame with a weight of $(1 - \lambda_t(i, j, k))$. Motion vector $(d_x(i, j), d_y(i, j))$ specifies the horizontal and vertical displacements at location $(i, j)$ in frame $k$ to its correspondence in frame $k - 1$. Bilinear interpolation is carried out at the non-integer locations on the temporal error image $\varepsilon_t$. The $\alpha_i$ are the spatial error diffusion filter coefficients with $\sum_i \alpha_i = 1$. In Fig. 2, the spatial neighbors $\mathcal{S}$ and $\alpha_i$

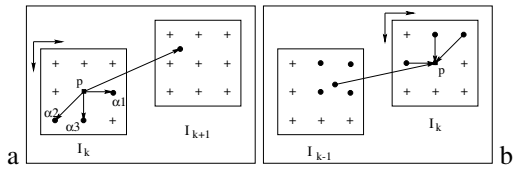


**Fig. 2**. Error diffusion of (a) $\varepsilon^+(p)$ and (b) $\varepsilon^-(p)$.

are chosen as those defined in the variable-coefficient error diffusion [5], with $\alpha_i$ varying with intensity code value $I(i, j, k)$ and $\mathcal{S} = \{(1, 0), (-1, 1), (0, 1)\}$.

The temporal diffusion map $\lambda_t(i, j)$ on frame $k$ (also denoted as $\lambda_t(i, j, k)$ in (3) is content dependent, and can be determined by the temporal characteristics of the HVS and the video frame rate. Based on the psychophysical experiments, the temporal model [6] consists of a lowpass filter and a bandpass filter. Specifically, it uses function $h_t(t) = \exp\{-(\frac{\ln(t/\tau)}{\sigma})^2\}$ and its high order derivatives to model the temporal mechanism of the targets perceived at the center of the human eye. Based on the temporal filters $h_t(t)$ and $h_t''(t)$, we choose $\lambda_t(i, j)$ as

$$\lambda_t(i,j,k) = 1 - \exp\{-\frac{(I(i,j,k) - \bar{I}(i,j,k))^2}{2\sigma_t^2}\}, \quad (4)$$

so that the major part of the noise energy falls into the stopbands. In (4), $\bar{I}(i, j, k) = h_t(k) \otimes I(i, j, k)$ is the temporally smoothed version of $I(i, j, k)$. At low frame rates ($< 10Hz$), $\lambda_t = 0$ as $\bar{I}(i, j, k) = I(i, j, k)$, there is no temporal error diffusion, and all of the quantization errors are exclusively diffused to spatial neighbors. It is the same situation in the static regions at high frame rates. In the fast-moving regions at high frame rates, $\lambda_t$ approaches 1, allowing more quantization error to diffuse across frames and leaving less errors to be diffused in the spatial domain. The high frequency noises become less visible after temporal smoothing by the HVS. At frame rates higher than 60 Hz, the temporal masking effect of the human eye should be taken into consideration. As the sensation of high contrast pattern lasts a finite duration, some frames can be dropped.

## 3. TEMPORAL CONSISTENCY

Temporal flicker is a special artifact that, over time, alternates black and white patterns at the same spatial location. It can be caused by model approximation or independent intraframe halftoning. To alleviate temporal flicker, we use adaptive gain control to increase the temporal consistency in $\mathbf{V}_d$ by adaptively changing the threshold

$$T(i,j,k) = (1 - \text{sign}\{I_d(i,j,k-1) - I_m\} \cdot \lambda_g(i,j,k)) \cdot I_m \quad (5)$$

used in the quantization decision (1). The adaptive gain control increases the inertia of interframe halftoning, making

| Temporal filter | WSNR (dB) for "Trevor" at 30 Hz Lowpass $h_{30}(t)$ | Bandpass $h''_{30}(t)$ |
|---|---|---|
| Vtone | 23.9247 | 3.3164 |
| Floyd-Steinberg | 27.5038 | -0.6500 |
| Ordered-dither | 21.2181 | -0.4660 |
| CG [2] | 24.5480 | 3.0875 |
| AFHBA [3] | 30.5215 | 0.8776 |

**Table 1**. Performance comparison.

$I_d(i,j,k)$ similar to $I_d(i,j,k-1)$ unless the spatiotemporally diffused error, $\varepsilon^-(i,j,k)$, is large enough.

The gain control map, $\lambda_g(i,j)$, on frame $k$ (also denoted as $\lambda_g(i,j,k)$) is content dependent and can be chosen as

$$\lambda_g(i,j) = \exp\{-\frac{d_x^2(i,j)+d_y^2(i,j)}{2\sigma_g^2}\}, \qquad (6)$$

where $(d_x, d_y)$ is the motion vector from point $(i,j)$ in frame $k$ to its correspondence in frame $k-1$. Numerous motion estimation algorithms can be used to compute $(d_x, d_y)$, such as gradient-based, region-based, energy-based, and transform-based approaches. For some compressed input video, such as the MPEG, QUICKTIME, or streaming video, the block motion vectors are readily available in the data stream without further computation. In static and slow-moving regions, $\lambda_g(i,j)$ is close to 1, and the halftoning of $I(i,j,k)$ is strongly biased to $I_d(i,j,k-1)$ for enhanced temporal consistency. In fast-moving regions with large motion vectors, $\lambda_g(i,j)$ is close to 0, and free error diffusion is encouraged to conceal the quantization error. Scale factor, $\sigma_g$ (e.g., 0.75), guides the transition from slow to fast motion.

## 4. EXPERIMENTAL RESULTS

Selected frames of the "Trevor" sequence ($256 \times 256$; 99 frames), with the motion fields to the previous frames, are shown in Fig. 3(a). The 8-bit grayscale sequence is rendered as a monochrome video with only black and white dots. Frames 34 of the halftone videos rendered at frame rates of 30 Hz and 60 Hz are printed in Fig. 3(b) at a spatial resolution of 120 dpi. The random patterns, coupled with the characteristics of HVS, provide a sensation of enhanced tone scale. The video halftoning algorithm is also applied to colortone video generation. Selected frames of the 24-bit color "Football" sequence ($360 \times 240$; 97 frames), with motion vectors, are shown in Fig. 4(a). The continuous tone color video is rendered as a colortone video with a palette of only 8 colors. The results on frame 34 at frame rates of 30 Hz and 60 Hz are printed in Fig. 4(b) at 120 dpi. The colortone video uses only a fraction of the colors to give a realistic tone scale rendering. Examples of the gain control maps, $\lambda_g(i,j)$, and the temporal diffusion map, $\lambda_t(i,j)$, on frame 34 of the sequences are shown in Fig. 5.

The video halftoning algorithm is compared to various halftoning techniques, including the Floyd-Steinberg error diffusion method, the ordered-dither method, the frame dependent image halftoning method (CG) [2], and the 3-D error diffusion method (AFHBA) [3]. The numeric results, in terms of weighted signal-to-noise ratio (WSNR) at frame rate of 30 Hz, are presented in Table 1, where WSNR is a measure of the spatiotemporally filtered signal energy over the spatiotemporally filtered noise energy defined as WSNR $=$

$$10\log\frac{\sum_{ijk}(h_s(i,j)\otimes h_t(k)\otimes I(i,j,k))^2}{\sum_{ijk}(h_s(i,j)\otimes h_t(k)\otimes (I(i,j,k)-I_d(i,j,k))^2}. \qquad (7)$$

Overall, the video halftoning (Vtone) technique provides the best spatiotemporal halftone rendering of the original grayscale continuous tone video.

## 5. CONCLUSIONS

We have presented a video halftoning algorithm to render continuous tone digital video as halftone and colortone sequences by the use of separable 1-D temporal and 2-D spatial error diffusions. A motion-adaptive gain control scheme is also proposed to enhance temporal consistency and alleviate flickering artifacts.

## 6. REFERENCES


[1] H. Hild and M. Pins, "A 3-D error diffusion dither algorithm for half-tone animation on bitmap screens," in *State-of-the-Art in Computer Animation – Proceedings of Computer Animation*, Geneva, 1989, pp. 181–190.

[2] C. Gotsman, "Halftoning of image sequence," *The Visual Computer*, vol. 9, no. 5, pp. 255–266, 1993.

[3] C. B. Atkins, T. J. Flohr, D. P. Hilgenberg, C. A. Bouman, and J. P. Allebach, "Model-based color image sequence quantization," in *Proc. of SPIE / IS&T Conf. on Human Vision, Visual Processing, and Digital Display V*, February 1994, vol. 2179, pp. 310–317.

[4] E. H. Adelson and J. R. Bergen, "Spatiotemporal energy models for the perception of motion," *Journal of Optical Society of America A*, vol. 2, no. 2, pp. 284–299, February 1985.

[5] V. Ostromoukhov, "A simple and efficient error-diffusion algorithm," in *Proceedings of ACM SIGGRAPH*, pp. 567–572, 2001.

[6] R. E. Fredericksen and R. F. Hess, "Estimating multiple temporal mechanisms in human vision," *Vision Research*, vol. 38, pp. 1023–1040, 1998.

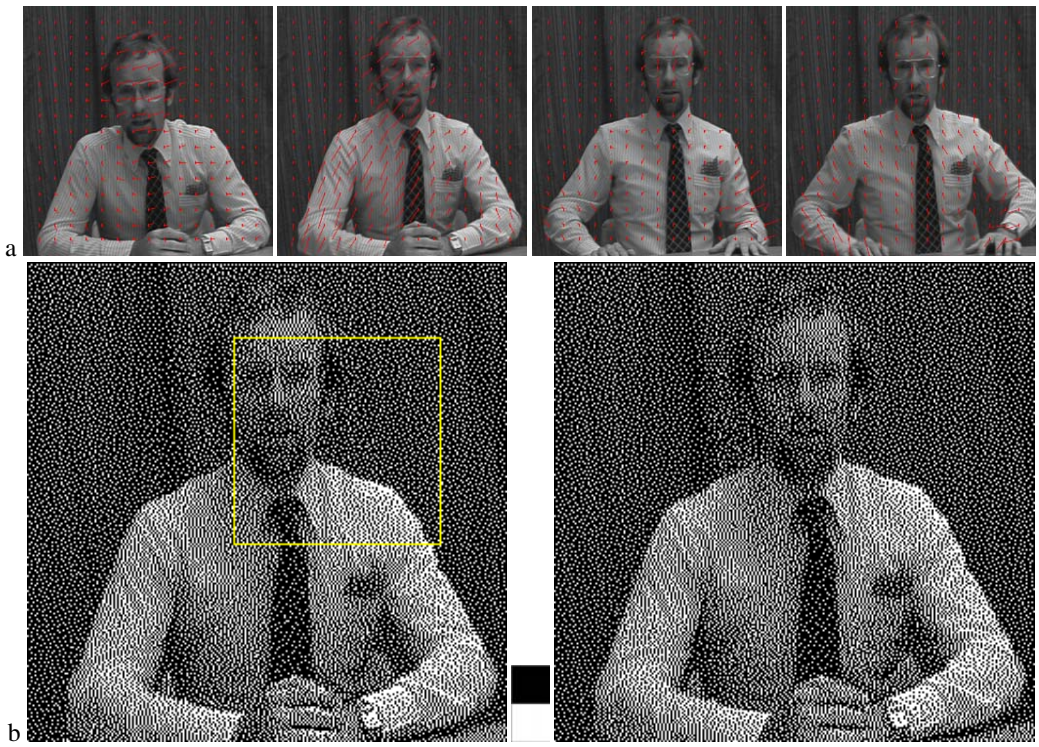


**Fig. 3**. (a) Grayscale video sequence "Trevor" with motion vectors. (b) Halftone video at 30 Hz (left) and 60 Hz (right).

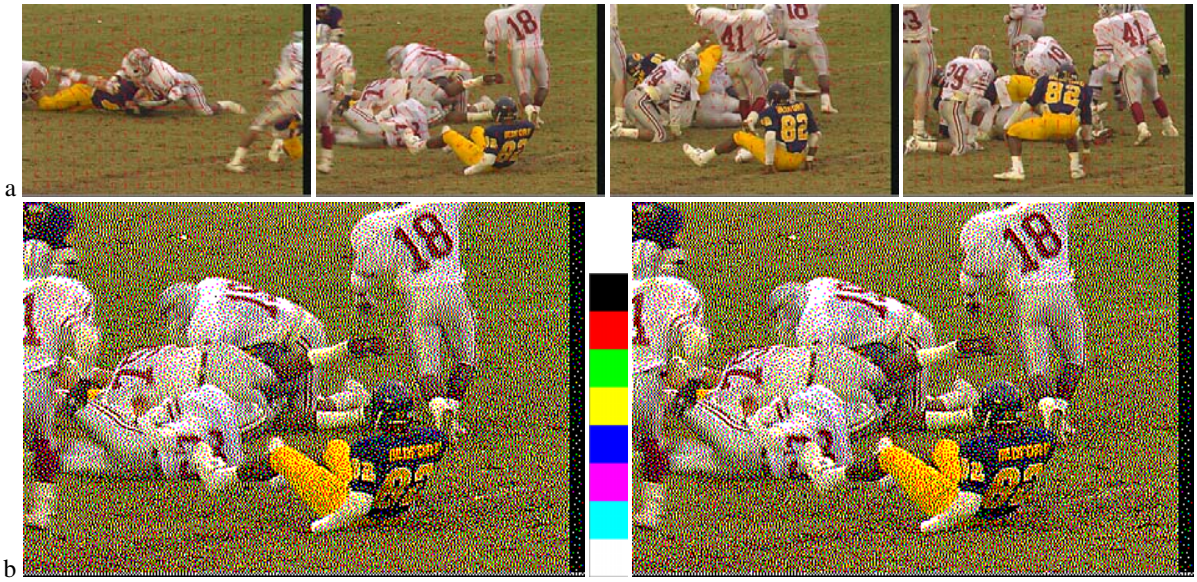


**Fig. 4**. (a) Color video sequence "Football" with motion vectors. (b) Colortone video at 30 Hz (left) and 60 Hz (right).

**Fig. 5**. Gain control maps (left) and temporal diffusion maps (right) of the "Trevor" and the "Football" sequences at 30 Hz.