Video halftoning

Zhaohui Sun

IEEE Transaction on Image Processing,
vol. 15, no. 3, pp. 678-686, March 2006

School of Electrical Engineering and Computer Science
Kyungpook National Univ.
Abstract

◆ Video halftoning
  – Rendering a digital video sequence onto display devices
    • Limited intensity resolutions and color palettes
    • Trading the spatiotemporal resolution for enhanced intensity/color resolution
    • Continuous tone video is not necessary and not practical for video display

◆ Proposed
  – Diffused quantization error of a pixel
    • Separable 1D temporal and 2D spatial error diffusions
  – Motion-adaptive gain control
    • To enhance the temporal consistency of the visual patterns by minimizing the flickering artifacts
1. Introduction

◆ Video halftoning
  – Display
    • Mismatch between image/video representation and display capability
    • Cellular phone, PDA, cinema poster, commercial billboard, etc.
  – Data reduction
    • Shorter bit depth, small size of data
  – Error-resilient communication
    • Channel noise is less pronounced in terms of image quality degradation
📍 Conventional approach
  – For minimizing the flickering artifacts
    • 3D error diffusion
    • Iterative image halftoning
      – Halftone map on the previous frame is used as the starting point for iterative refinement on the current image frame
    • Spatiotemporal error diffusion filters
      – Luminance and chrominance channels at different frame rates
    • Direct-binary-search algorithm
      – Applied to 3D error diffusion
◆ Major contribution
  – Scheme of video tone-scale reduction by separable temporal and spatial error diffusion
  – Method of temporal flicker reduction by the use of motion information
2. Problem formulation

- Digital video sequence
  
  \[ V = \{ I(i, j, k), i = 1...M, j = 1,...N, k = 1,...K \} \]
  
  - Single luminance channel and two chrominance channels
  - Perceived visual difference

  \[ \varepsilon(i, j, k) = h_e(i, j, k) \otimes (V(i, j, k) - V_d(i, j, k)) \]  
  
  Fig. 1. Video halftoning finds the best halftone rendering \( V_d \) with the minimal perceived visual difference \( \varepsilon \).

\( V \)

Halftoning

\( V_d \)

Display Device

Visual System

\( \varepsilon \)
Separable in temporal and spatial

\[ \varepsilon(i, j, k) = \sum_{i', j'} h_s(i - i', j - j') \cdot \left[ \sum_{k'} h_t(k - k') \cdot \left( I(i', j', k') - I_d(i', j', k') \right) \right] \] (2)

Formulated as an optimization problem

\[ \mathbf{V}_d^* = \arg \min_{\mathbf{V}_d} \| h_e \otimes (\mathbf{V} - \mathbf{V}_d) \|^2 \] (3)

Intensity values are normalized [0,1]
3. Spatiotemporal error diffusion

◆ Three dimensional error diffusion
  – Spatiotemporal model of the HVS (Modulation Transfer Function)

\[
H_e(f_s, f_t) = \left\{ 6.1 + 7.3 \ln \left( \frac{f_t}{3f_s} \right)^3 \right\} (2\pi)^2 f_s f_t \times \exp \left\{- \frac{4\pi(f_t + 2f_s)}{45.9} \right\}
\] (4)

where \( f_s \) is the spatio-temporal frequency in cycles per degree
\( f_t \) is the temporal frequency in Hz

- Lowpass in spatial dimension and bandpass in temporal dimension
- Spreading the quantization error to the stop bands as high frequency noise (blue noise) such that they are less visible
– Optimal halftone video $V_d$

incoming video $V$, diffused error $\hat{V}_e \Rightarrow \hat{V} = V + \hat{V}_e$
halftone video $V_d$, quantized error $V_e \Rightarrow \hat{V}_e = \hat{V} - V_d$
spatiotemporally filtered quantized error $\hat{V}_e = h_e \otimes V_e$

– Requiring large memory and intensive computation
– 3D error diffusion has a few difficulties in practice
  • Practiced only local optimal solution
  • Intensive computation, delay, high system complexity
    – Introducing additional artifacts, such as temporal flicker
  • Separable temporal and spatial filters
◆ Separable error diffusion
  
  – Approximated MTF \(H_e(f_x, f_y, f_t)\)
    
    • 1D temporal MTF and 2D spatial MTF
      \[H_e(f_x, f_y, f_t) = H_s(f_x, f_y) \cdot H_t(f_t)\]
  
  – Basic idea
    
    • Diffusing
      
      – Part of the quantization error of a pixel into its causal temporal neighbor along the motion trajectory
      – Remainder to its causal spatial neighbors
    
    • Controlled exact amount by temporal diffusion map
      
      – In fast-moving regions, more interframe error is diffused temporally and leaving less error to be diffused to the spatial neighbors
Separable error diffusion scheme

Fig. 2. Separable temporal and spatial error diffusion with motion-adaptive gain control.
- Quantized intensity

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

(5)

- Quantized error

\[ \varepsilon^+(i, j, k) = \hat{I}(i, j, k) - I_d(i, j, k) \]  

(6)
– Diffused error

• Separable error to temporal and spatial domain

\[
\varepsilon^{-}(i, j, k) = \lambda_t(i, j, k) \cdot \varepsilon_s(i + d_x(i, j), j + d_y(i, j), k - 1) + (1 - \lambda_t(i, j, k)) \cdot \sum_{s' \in S} \alpha_i(I(i, j, k)) \cdot \varepsilon_s(i + s^j_x, j + s^j_y, k)
\]  

(7)

Fig. 3. (a) Diffusion of error \(\varepsilon^{+}(p)\) to its spatial neighbors and the temporal neighbor on the next frame and (b) collection of error \(\varepsilon^{-}(p)\) from its spatial neighbors \(\varepsilon_s\) and temporal neighbor \(\varepsilon_t\).
• Motion vector, \((d_x(i,j), d_y(i,j))\)
  - Specifying the horizontal and vertical displacement in frame \(k\)
  - Bilinear interpolation is carried out at the noninteger locations

• Spatial error diffusion filter coefficients, \(\alpha_i\)

\[
\sum_i \alpha_i = 1 \quad \text{and} \quad S = \{(1,0), (-1,1), (0,1)\}
\]

  - Filter coefficients are chosen as those defined in the variable-coefficient error diffusion, with coefficient varying with intensity
Temporal diffusion

- Temporal characteristics of the HVS
  - Consist of a lowpass filter and a bandpass filter

\[
g_t(t) = \exp \left\{ - \left( \frac{\ln(t_0)}{\sigma} \right)^2 \right\}
\]

(8)

- Function \( g_t(t) \) and its normalized second-order derivative \( g''_t(t) \)

Fig. 4. Temporal characteristics of the human visual system (a) Impulse response and (b) frequency response.
• Approximation of $g_t(t)$ and $g''_t(t)$ by FIR filter design
  – Five-tap filter for 30Hz
  – Nine-tap filter for 60Hz

Fig. 5. (a) Lowpass filter $g_{30}(t)$ and the bandpass filter $g''_{30}(t)$ at 30Hz (b) Lowpass filter $g_{60}(t)$ and the bandpass filter $g''_{60}(t)$ at 60Hz.
- Temporal diffusion map, $\lambda_t(i, j)$
  - Content dependent
  - Determined by the temporal characteristics of the HVS and the video frame rate

$$
\lambda_t(i, j, k) = 1 - \exp \left\{ - \frac{(I(i, j, k) - \bar{I}(i, j, k))^2}{2\sigma_t^2} \right\} \quad (9)
$$

where $\bar{I}(i, j, k) = g_t(k) \otimes I(i, j, k)$ is temporally smoothed version of $I(i, j, k)$

- Major part of the noise energy falls into the stop bands
- At low frame rate (<10Hz)
  - No temporal diffusion, $\lambda_t = 0$ as $\bar{I}(i, j, k) = I(i, j, k)$
- At high frame rate
  - more temporal diffusion, $\lambda_t$ approaches 1
Spatial diffusion

- Spatial frequency response of the HVS
  \[ H_s(f_s) = 2.6(0.0192 + 0.114 f_s) \exp\{-(0.114 f_s)^{1.1}\} \tag{10} \]
  where \( f_s = \sqrt{(f_x^2 + f_y^2)} \) : frequency in degrees per cycle

- Low pass characteristics
  - a peak at 8 cycles/degree
  - Dropping to 0 beyond 30 cycles/degree

- Distributed quantization error to the high frequency bands as the less visible, blue noise.
◆ Color video
  – Extension of separable error diffusion
    • Multiple tone-scale levels
      – Multilevel display
      – Replaced the binary thresholding by multilevel quantizer
    • Multiple color channels
      – If Color dependency is ignored, simply applied to each channel
    • Desired to explore the color dependency
4. Motion-adaptive temporal consistency

- **Temporal flicker**
  - Alternating black and white patterns at the same spatial location
    - Caused by model approximation or independent intraframe halftoning
  - Adaptive gain control
    - to increase the temporal consistency in $V_d$ by adaptively changing the threshold

\[
T(i, j, k) = \begin{cases} 
(1 - \lambda_g(i, j, k))I_m, & \text{if } I_d(i, j, k - 1) \geq I_m \\
(1 + \lambda_g(i, j, k))I_m, & \text{otherwise} 
\end{cases} 
\]  

(11)

where $I_m$ is the middle point
– Gain control map $\lambda_g(i, j)$ on frame $k$

$$\lambda_g(i, j) = \exp\left\{-\frac{(d_x^2(i, j) + d_y^2(i, j))}{2\sigma_g^2}\right\} \quad (12)$$

where $(d_x, d_y)$ is the motion vector from point $(i, j)$ in frame $k$

• In slow moving region, $\lambda_g(i, j)$ is closed to 1
• In fast moving region, $\lambda_g(i, j)$ is closed to 0

– Motion estimation

• Numerous algorithm, such as gradient-based, region-based, energy-based, and transform-based approaches
• Block motion vectors
– Using temporal variances of adjacent frames instead of the motion vectors
  • Windowed average of temporal intensity

\[ \lambda_g(i, j) = \exp \left\{ - \frac{\text{E}\{I(i, j, k) - \text{E}\{I(i, j, k)\}\}^2}{(2\sigma_g'^2)} \right\} \]  

(13)

where espection \( \text{E}\{I(i, j, k)\} = (1/(2q + 1)) \sum_{k' = -q}^{q} I(i, j, k + k') \)

– Using the temporal highpass filtering as a measure of the intensity changes

\[ \lambda_g(i, j) = \exp \left\{ - \frac{(h_h(k) \otimes I(i, j, k))^2}{(2\sigma_g''^2)} \right\} \]  

(14)

where \( h_h(k) \) is a bandpass/highpass temporal filter
5. Algorithm

Input: Digital continuous tone video
\[ V = \{I(i,j,k)\} \]

Output: Halftone/color tone video \[ V_{df} = \{I_{df}(i,j,k)\} \]

1) Initialize temporal filter \( h_t \), temporal diffusion map \( \lambda_t(i,j) = 0 \), gain control map \( \lambda_g(i,j) = 0 \), motion field \( (d_x, d_y) = (0,0) \), and frame index \( k = 1 \).

2) Process the frames sequentially, and scan the pixel \( p = (i,j,k) \) on frame \( k \) in a serpentine order.

3) Collect the diffused quantization error \( \varepsilon^{+}(i,j,k) \) from the spatiotemporal neighbors (7).

4) Quantize \( I(i,j,k) \) to \( I_{df}(i,j,k) \) based on the motion-adaptive quantization threshold (5).

5) Compute the quantization error \( \varepsilon^{+}(i,j,k) \) (5).

6) Diffuse part of \( \varepsilon^{+}(i,j,k) \) to the temporal neighbor along the motion trajectory.

7) Diffuse the rest of \( \varepsilon^{+}(i,j,k) \) to the causal spatial neighbors on frame \( k \).

8) Go to step 2) for the rest of the pixels on frame \( k \), then increase frame index \( k \leftarrow k + 1 \).

9) Retrieve or compute the motion field \( (d_x, d_y) \) from frame \( k \) to frame \( k - 1 \).

10) Determine the temporal diffusion map \( \lambda_t(i,j) \) (9).

11) Determine the gain control map \( \lambda_g(i,j) \) (12), (13), (14).

12) Go to step 2) until all the frames are processed.
6. Experimental results

◆ Setup
  
  – Tested on two video sequences
    
    • Grayscale “Trevor” sequence
      – 99 frames and spatial resolution of 256x256
    
    • Color “Football” sequence
      – 97 frames and spatial resolution of 360x240
  
  – Weighted signal-to-noise ratio (WSNR) for evaluation

\[
WSNR = 10 \log \left( \frac{\sum_{ijk} (h_s(i,j) \otimes h_t(k) \otimes I(i,j,k))^2}{\sum_{ijk} (h_s(i,j) \otimes h_t(k) \otimes (I(i,j,k) - I_d(i,j,k)))^2} \right)
\]  

where the temporal filter \( h_t(k) \) is chosen as \( g_t(k) \) or \( g_t''(k) \)
Halftone/colortone video
  – “Trevor” sequence

Fig. 6. (a) Frame 2, 34, 66, and 99 of the grayscale video sequence “Trevor” overlaid with motion vectors to the previous frames (b) Frame 34 of the halftone video at the frame rate of (left) 30Hz and (right) 60Hz.
Fig. 7. (a) Frame 2, 34, 66, and 97 of the color video sequence “Football” overlaid with motion vectors to the previous frames (b) Frame 34 of the colortone video at the frame rate of (left) 30Hz and (right) 60Hz.
Examples of gain control map and temporal diffusion maps on frame 34

- In static and slow-moving patterns, strongly biased to halftone patterns on the previous frames
- In fast-moving patterns, free error diffusion allowed for the best possible image reproduction

Fig. 8. Gain control map $\lambda_g(i, j)$ in (a) and (c) and temporal diffusion maps $\lambda_t(i, j)$ in (b) and (d) on frame 34 of the sequences at 30Hz
Evaluation

- Compared with five different halftoning techniques
  - 30Hz, “Trevor” sequence, frame 34

Table 1. Performance comparison of the halftoning schemes.

<table>
<thead>
<tr>
<th>Temporal Filter</th>
<th>WSNR (dB) for “Trevor” at 30 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lowpass $g_{30}(t)$</td>
</tr>
<tr>
<td>Video halftoning</td>
<td>23.9247</td>
</tr>
<tr>
<td>Floyd-Steinberg [2]</td>
<td>27.5038</td>
</tr>
<tr>
<td>Ordered dither [3]</td>
<td>21.2181</td>
</tr>
<tr>
<td>AFHBA [14]</td>
<td>30.5215</td>
</tr>
</tbody>
</table>
– Halftoned frame and flickering artifact at 30Hz

Fig. 9. Halftone frame 34 and the flickering artifact at 30 Hz by six methods.
7. Conclusion

- **Video halftoning scheme**
  - To render continuous tone digital video on display device with limited tone scales and color palettes
  - Separable 1D temporal and 2D spatial error diffusions
  - Temporal FIR filter are designed at various video frame rates
    - To diffuse temporal errors along motion trajectory across frames
  - Motion-adaptive gain control scheme
    - To enhance temporal consistency and alleviate flickering artifacts