ECE 418: Introduction to Image & Video Processing

Video scanning and display

Shuai Huang

02/28/2019

University of Illinois at Urbana-Champaign
Overview

1. Time-varying imagery: video
   1.1 Interlaced scan
   1.2 High-definition television
Time-varying imagery: video
Time-varying imagery: video

The eyes see time-varying imageries:

- 3D signal: two spatial and one temporal coordinates.

Digital processing system takes an image sequence at a certain rate, know as video.

- Motion picture: 48/60/72 images/sec.
- Television: 60 images/sec in North America and Japan, 50 images/sec in Europe.
- Computer monitor: typically 60 images/sec. 120/240 images/sec for high-end gaming monitors.
- Cell phone: 60 images/sec.
Time-varying imagery: video

The quality of the video depends on

- Spatial resolution.
- Temporal resolution.

We focus on the temporal properties of video in this lecture.
Temporal properties of video

Temporal summation

Temporal summation is the eye’s ability to sum/integrate light’s effect over time.

- Only occurs within a certain period of time $t \leq t_c$, where $t_c$ is the critical duration.
- Temporal summation stops once a threshold $k = L \times t_c$ is reached (Bloch’s Law).

The critical duration $t_c$ depends on the luminance of the light $L$.

- When $L$ is large, $t_c$ is small.
- When $L$ is small, $t_c$ is large.
Temporal properties of video

The detection of light occurs at the retina.

- The integration time of **cones** is around 10~15 ms.
- The integration time of **rods** is around 100 ms.

For bright light, cones take over.

- Smaller integration time = higher temporal resolution.
Figure 1: Flash of light presented to the eye: (a) With a short integration time, the flashes are detected; (b) With a long integration time, no flashes are perceived.
Flicker

- The flickering is detected when you notice the change in light intensity within a short period of time.

Consider the following spatially-uniform signal that varies in time with frequency $f_T$

$$I(x, y, t) = L + \Delta L \cos(2\pi f_T t)$$  \hspace{1cm} (1)

- When $f_T$ is large enough, $I(x, y, t)$ cannot be distinguished from the constant image

$$I(x, y, t) = I.$$  \hspace{1cm} (2)
Critical Flicker Fusion Frequency

The frequency $f^*_T$ when the time-varying signal $I(x, y, t)$ cannot be distinguished from a constant image by the human eye.

- When $f_T > f^*_T$, $I(x, y, t) = L + \Delta L \cos(2\pi f_T t) = L$ to the human eye.
- $f^*_T$ depends on the magnitude of variation $\Delta L$. 

Flicker
Flicker

For every frequency $f_T$, when the variation magnitude $\Delta L$ is below some threshold $\tau_L$, the eye does not perceive the flicker.

- The threshold $\tau_L$ also depends on the background luminance $L$.
- For all $L$, the threshold $\tau_L$ increase rapidly when $f_T > 30$ Hz.

In summary, $\tau_L$ depends on $f_T$ and $L$.

Contrast sensitivity

The following ratio is the contrast sensitivity.

$$\frac{L}{\tau_L}$$ (3)
Figure 2: The horizontal axis shows the frequency $f_T$, the vertical axis shows the threshold $\tau_L$, different lines correspond to different background luminance $L$.
Flicker

- Above the line, there is no flicker.
- Below the line, flicker can be detected.

### Flicker detection

The eye is more sensitive to flicker at high background luminance levels than at low luminance levels.

- Computer/TV screens are bright, needs a high refresh rate 60 images/sec.
- The background luminance in a movie theater is low, 48 images/sec is acceptable: the same image is usually displayed twice to avoid flickering.
Spatio-temporal modulation transfer function

Review: MTF represents the contrast sensitivity as a function of the spatial frequency for still images.

Figure 3: Modulation transfer function of the human eye.

Extend MTF to the spatio-temporal frequencies for videos.
Spatio-temporal modulation transfer function

The test image is varying spatially and temporally.

\[ I(x, y, t) = L + \Delta L \cos(2\pi f_S x) \cos(2\pi f_T t) \]  \hspace{1cm} (4)

Question:

○ Where is \( y \)?

○ What does this image look like?

Given \( f_S \) and \( f_T \), there is a threshold \( \Delta L < \tau_L \) under which flickering is not visible. The contrast sensitivity is

\[ \frac{L}{\tau_L} \]  \hspace{1cm} (5)
Spatio-temporal modulation transfer function

Figure 4: Spatio-temporal MTF.
Figure 5: Cross-sections of the spatio-temporal MTF, for fixed temporal frequencies.
Figure 6: Cross-sections of the spatio-temporal MTF, for fixed spatial frequencies.
Spatio-temporal modulation transfer function

Important conclusions

- The eye has poor contrast sensitivity at high spatio-temporal frequencies.
- Flicker is more visible for low-spatial-frequency patterns than for high-frequency ones.

For example, flicker is more commonly seen for a screen showing a uniform color, i.e. low spatial frequency.
### Motion perception

The perception of a moving object is more or less accurate depending on whether the eye tracks the object or not.

- It is in general **impossible** to predict whether the viewer’s eye will track objects.

### Video fidelity criterion

The still-image fidelity criterion can be extended to the temporal domain: a common choice is the mean square error (MSE)

- There is no universally good criterion.
- In particular, it is difficult to quantify motion perception.
Sampling and reconstruction

We mostly work with digitized 3D discrete signal.

**Sampling**

When sampling is done well below the temporal Nyquist rate, considerable aliasing occurs.

- Hovering helicopter.
- Car wheels spinning backwards.

**Reconstruction**

Reconstruction of a continuous signal from its samples requires an interpolation filter.

- Human eye naturally performs the interpolation.
- Computer can do it too, such as the SmoothVideo Project (SVP).
Motion pictures

Typical movie cameras take 24 images/sec, and record them in analog or digital form.

- Each image is a shot of a scene, 24 shots are taken within 1 second.
- Each image is exposed for a period of $T_{ex}$. Note that $T_{ex} < \frac{1}{24}$. 
Typical digital video can be modeled as

\[ x(n_1, n_2, n_3) = y_a(n_1T_1, n_2T_2, n_3T_3) \]

\[ = \int \int \int x_a(t_1, t_2, t_3) \cdot w(t_1 - u_1, t_2 - u_2, t_3 - u_3) \, dt_1 \, dt_2 \, dt_3 , \]

(6)

where the 3D sampling aperture remain the same within the exposure time of a single shot, and zero otherwise.

\[ w(t_1, t_2, t_3) = \begin{cases} 
  w_{12}(t_1, t_2) & \text{if } 0 \leq t_3 \leq T_{ex} \\
  0 & \text{otherwise} 
\end{cases} \]

(7)
Size of digitized movies

Figure 7: 35mm format.
Size of digitized movies

- Assume a resolution of 100 lines/mm.
- The film should be sampled at 200 samples/mm.

For an image taken with the usual 35mm format (or 135 format) film, i.e. 36 × 24 mm, with a 16 × 22 mm active area (the academy format, 1.375:1), we obtain 3200 × 4400 pixels for each color component.

- Using 10 bit per pixel and 24 frames/sec, the bit rate is

\[ 24 \times 10 \times 3 \times 3200 \times 4400 \approx 10 \text{Gbit/s} \]  

- Even for nowadays, this is too much data!
Size of digitized movies

The actual bit rate is significantly lower,

- The actual spatial and amplitude resolutions are smaller than quoted previously.
- Digital compression is used.
Interlaced scan

Interlacing doubles the perceived frame rate of a video display.

- The interlaced signal contains two fields of a video frame.
Figure 8: Comparison of progressive and interlaced scans.
Interlaced scan

- Low spatial/temporal frequencies are reproduced well.
- High spatial/temporal frequencies are severely distorted.

Figure 9: Interlacing artifacts.
Interlaced scan

The obvious advantage:

- Twice as economical as the progressive scan in terms of number of lines/sec.

The potential and less convincing advantage:

- Aliasing of high spatio-temporal frequency is often acceptable because frequency components may be filtered out by the HVS.
High-definition television

HDTV in various formats

- 720p (HD ready): 1280×720p: ~ 0.92 MP per frame.
- 1080i (full HD): 1920×1080i: ~ 1.04 MP per field or ~ 2.07 MP per frame.
- 1080p (full HD): 1920×1080p: ~ 2.07 MP per frame.

“p” stands for progressive scan, “i” stands for interlaced scan.

- 8 to 15 Mbit/s typ - HDTV quality.
High-definition television

Figure 10: Worldwide HDTV standards.
ATSC standard

ATSC is adopted in US on 06/12/2009.

Audio

Dolby Digital AC-3 is used as the audio codec.

- Support five channels of sound with a sixth channel for low-frequency effects (5.1 configuration).

Video

MPEG-2 video system is used to transport data.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Aspect ratio</th>
<th>Pixel aspect ratio</th>
<th>Scanning</th>
<th>Frame rate (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical</td>
<td>Horizontal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1080</td>
<td>1920</td>
<td>16:9</td>
<td>1:1</td>
<td>progressive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23.976</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>interlaced</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29.97 (59.94 fields/s)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30 (60 fields/s)</td>
</tr>
<tr>
<td>720</td>
<td>1280</td>
<td>16:9</td>
<td>1:1</td>
<td>progressive</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>23.976</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>29.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>59.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>60</td>
</tr>
</tbody>
</table>
### ATSC Standard A/53 Part 4:2009 (MPEG-2 Video System Characteristics)

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Aspect ratio</th>
<th>Pixel aspect ratio</th>
<th>Scanning</th>
<th>Frame rate (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical</td>
<td>Horizontal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>480</td>
<td>704 or 858</td>
<td>4:3 or 16:9</td>
<td>SMPTE 259M</td>
<td>23.976, 24, 29.97, 30, 59.94, 60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>progressive</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>interlaced</td>
<td>29.97 (59.94 fields/s), 30 (60 fields/s)</td>
</tr>
<tr>
<td>640</td>
<td>4:3</td>
<td>1:1</td>
<td>progressive</td>
<td>23.976, 24, 29.97, 30, 59.94, 60</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>interlaced</td>
<td>29.97 (59.94 fields/s), 30 (60 fields/s)</td>
</tr>
</tbody>
</table>
ATSC 3.0

ATSC 3.0 is the next major version to be widely deployed.

- Support video channels up to 2160p 4K resolution at 120 frames per second.
- Wider color gamut, HDR, Dolby AC-4, etc.

The first major deployment of ATSC 3.0 occurred in South Korea in 2017 for the 2018 Winter Olympics.
What would be the bit rate of a system that is limited by the HVS’s spatiotemporal resolution?

- The number of photoreceptors (rods and cones) in each eye: \( \sim 160,000,000 \)
- The response of each photoreceptor is encoded using less than 10 bits.
- The temporal resolution of the eye is assumed to be limited by 60 frames/sec.

\[ 2 \times 60 \times 10 \times 160,000,000 \approx 192 \text{Gbit/s} \]
The bit rate of IMAX format could reach as much as 150 Gbit/s.

IMAX is a system of

- High-resolution cameras.
- Film formats.
- Film projectors.
IMAX

Cameras are specially designed to shoot high-resolution videos in analog or digital format.

- Approximately 12,000 lines of horizontal resolution (12K).
The IMAX format is generally referred to as “15/70” film. It is much larger compared to the standard 35mm academy format.
The projection is performed via digital projection or laser projection.

- Usually 4K resolution nowadays.
- The standard screen is 22 m × 16.1 m (72 ft × 53 ft), but can be significantly larger.
Acknowledgment: The slides are based on the lecture notes by Prof. Pierre Moulin, UIUC.