TUTORIAL ON HIGH DYNAMIC RANGE VIDEO

Erik Reinhard
Giuseppe Valenzise
Frederic Dufaux
About the presenters

- **Erik Reinhard: Distinguished Scientist at Technicolor R&I**
  - Academic positions at universities and research institutes in Europe and North America
  - Founder and editor-in-chief of ACM Transactions on Applied Perception
  - Author of a reference book on HDR

- **Giuseppe Valenzise: CNRS researcher at LTCI, Telecom ParisTech**
  - Ph.D. in Information Technology at Politecnico di Milano, Italy
  - Experience in video coding, quality assessment and perceptual studies

- **Frederic Dufaux: CNRS Research Director at LTCI, Telecom ParisTech**
  - Editor-in-Chief of Signal Processing: Image Communication
  - Participated in MPEG and JPEG committees
  - Research positions at EPFL, Emitall Surveillance, Genimedia, Compaq, Digital Equipment, MIT
Outline

- **Introduction**
  - Background and fundamental concepts related to HDR.
- **Perception**
  - Properties of human visual system
- **Capture/Acquisition**
  - Camera technologies, multi-exposure techniques, ghost removal
- **Tone Reproduction**
  - How to display HDR video on a standard display, curve-based solution, spatial processing, video processing, inverse tone reproduction
- **Color Management**
  - Gamut boundary management techniques, luminance - chroma interactions
- **Video Coding**
  - Compression techniques for HDR video, pre- and post-processing, on-going standardization activities in MPEG related to HDR and Wide Color Gamut
- **Display**
  - Display hardware, display characterization, dual modulation
- **Quality of Experience**
  - The concept of QoE in the context of HDR, objective measures to predict HDR image and video quality
- **Applications & Wrap up**
INTRODUCTION
Context and Trends

- **Higher spatial and temporal resolutions**
  - Ultra High Definition (UHD), 4K, 8K
  - High Frame Rate (HFR)
- **Higher pixel depth**
  - up to 14 bits per component
  - High Dynamic Range (HDR)
- **More colors**
  - 4:4:4 color sampling
  - Wide Color Gamut (WCG)
- **More views**
  - 3D, multi-view, free viewpoint
Context and Trends

- **Driven by**
- **Improved user experience**
  - More realistic scene rendering
  - More details
  - Immersive
  - Better perceived depth
- **New video technologies**
  - New devices
  - New multimedia services
Some definitions

- **Luminance**
  - A photometric measure of the luminous intensity per unit area
  - Candela per square meter, cd/m² (also referred to as *nits*)
Some definitions

- **Dynamic range or contrast ratio**
  - The ratio between the brightest and the darkest objects in a scene

- **f-stops**
  - Ratios of light or exposure
  - Defined in power-of-2 steps

- **Orders of magnitude**

\[
\frac{L_{\text{max}}}{L_{\text{min}}}
\]

\[
\log_2 L_{\text{max}} - \log_2 L_{\text{min}}
\]

\[
\log_{10} L_{\text{max}} - \log_{10} L_{\text{min}}
\]
Some definitions

• Low (Standard) Dynamic Range
  – ≤ 10 f-stops
  – Three channels, 8 bit/channel, integer
  – Gamma correction

• Enhanced (Extended, Wide) Dynamic Range
  – 10 to 16 f-stops
  – More than 8 bit/channel, integer

• High Dynamic Range
  – ≥ 16 f-stops
  – Floating points
  – Linear encoding
Human Visual System - Retina

- **Rods**
  - 100+ millions photoreceptor cells
  - Peripheral retina
  - Scotopic monochrome vision: $10^{-6}$ to $10$ cd/m$^2$

- **Cones**
  - 6 millions photoreceptor cells
  - Near the fovea
  - Photopic color vision: $0.03$ to $10^8$ cd/m$^2$
  - Long- (L, red), Medium- (M, green), and Short-wavelength (S, blue)
  - High resolution vision

Adapted after Østerberg, 1935
Human Visual System can adapt to a very large range of light intensities

- At a given time: 4-5 orders of magnitude
- With adaptation: 14 orders of magnitude
  - Mechanical, photochemical and neuronal adaptive processes
HDR Pipeline

Capture

Exposure stack

HDR camera

LDR display

HDR display

HDR image

Tonemapping
 HDR Pipeline

### Capture
- Bracketing
- HDR camera
- CG content

### Processing
- Merging/Fusion
- Tonemapping
- Color correction

### Distribution
- Format
- Encoding
- Display side TMO/iTMO

### Rendering
- Decoding

### Display
- LDR display
- HDR display

Tutorial on HDR Video, EUSIPCO 2016
New Challenges at each stage

• **Content creation**
  – True HDR cameras (when?)
  – Upgrading legacy content to HDR

• **Distribution**
  – Efficient video coding solutions
  – New standards for interoperability
  – Backward compatibility

• **Processing / rendering**
  – Tone mapping and inverse tone mapping
  – Color management

• **Display**
  – Professional displays up to 4000 nits
  – Consumer displays up to 1000 nits

• **Quality of experience**
  – Taking into account HVS properties
Capturing HDR video

F3 camera, 1080p 25 fps, 21 f-stops
When does HDR help most?

- Scenes with difficult lighting conditions
- Better opportunity to convey director’s intent
PERCEPTUAL PHENOMENA IN HDR
HDR and visual adaptation

- Human eye can perceive a range of approximately $14 \log_{10}$ units (46 f-stops)
- But only about $4 \log_{10}$ units at the same time
- High Dynamic Range perception through *adaptation*:  
  - Steady-state dynamic range (SSDR) compression  
  - Light/dark adaptation (rods and cones activation)  
  - Chromatic adaptation (color constancy)  

[Image: Diagram showing luminance range and visual adaptation process]
Steady-state Dynamic Range

- Between 3 and 4 \( \log_{10} \) units
- SSDR adaptation takes less than 500 ms
- Nonlinear response of photoreceptors (Naka and Rushton, 1966; Michaelis and Menten, 1913):

\[
\frac{V}{V_{max}} = \frac{L^n}{L^n + \sigma^n}
\]

where:
- \( V \) = signal response
- \( L \) = input luminance
- \( \sigma \) = semisaturation constant
- \( n = 0.7 \div 1 \)

Detectability based on cone response

\[ \Delta V_1 > \Delta V_2 \]

\[ \Delta L_1 = \Delta L_2 \]

[Kunkel et al. 2016]
Light/Dark adaptation

- The gain of all photoreceptors is adjusted to the background luminance
  - Similar to automatic exposure control in digital cameras
- The semisaturation constant is equal to the adaptation luminance
  - This process is local in the retina

Function of time:
- Full light adaptation takes about 5 minutes
- Full dark adaptation is much slower, and takes up to 30 minutes
- Time course of dark adaptation:

![Graph showing light adaptation and dark adaptation](image-url)
Chromatic adaptation

- The sensitivity of each cone photoreceptor (L, M, S) is adjusted *individually*
- Basic mechanism of color constancy
  - Similar to automatic white balance in digital cameras

[Fairchild, 2013]
Perception of luminance

• **Just-noticeable difference (JND):**
  - the minimum amount that must be changed in order for a difference to be detectable at least 50% of the times
  - also called *contrast detection threshold*
  - Contrast sensitivity = 1/threshold

• **Weber’s law (1834):** the JND between two stimuli is proportional to the magnitude of the stimuli, i.e.,
  \[
  \Delta R \sim \frac{\Delta L}{L}
  \]
  - \(\Delta L/L\) is called Weber ratio, and is assumed constant (about 1%)
  - when \(\Delta \to 0\), by integration, \(R \sim \log L\) (Fechner’s law, 1860)
  - valid only on a limited range of luminance

• **Stevens’ power law (1961):** \(R \sim L^\alpha\)
  - E.g., \(\alpha = 1/3\) in CIELAB brightness predictor
Spatial properties of the HVS

- The contrast detection threshold depends on the spatial frequency of a stimulus
- Campbell-Robson chart
  - Sinusoidal grating with exponentially increasing frequency and contrast
  - Michelson contrast measure: \( \frac{L_{\text{max}} - L_{\text{min}}}{L_{\text{max}} + L_{\text{min}}} \)
  - The contrast sensitivity function (CSF) is band-pass (low-pass for color)
CSF as a function of average luminance

- The CSF changes with the adaptation luminance
  - The contrast sensitivity increases with adaptation luminance, and drops in dimmer conditions where rods predominate and visual acuity is lower
- The peak of the CSF at a given luminance level gives the lowest detection threshold at that luminance
  - Conservative estimation of JND

[Aydin et al., 2008]
Contrast-versus-intensity (c.v.i.) function

- Describes how the detection threshold changes with adaptation luminance
  - Threshold-versus-intensity (t.v.i.) are computed on a single frequency
  - CSF-based models track the peaks of CSF’s (more conservative)

Weber’s law

[Weber’s law

Mantiuk et al., 2006]
Barten’s CSF model

- **Accurate closed-form model describing CSF for different adaptation luminances** (Barten, 1996)
  - Considers neural noise, lateral inhibition, photon noise, external noise, optical modulation transfer function, temporal filtering, etc.

- **Used in DICOM biomedical standard and for PQ / ST2084 EOTF**
  - Detection thresholds obtained at a fixed frequency (4 cpd) for DICOM
  - Detection thresholds obtained by tracking CSF peaks for PQ EOTF
Luminance to Luma mapping

- A luminance-to-brightness transfer function can be obtained as the cumulative sum of threshold values
  - Quantized with $2^n$ levels to obtain a luma signal

- Conventional 1/2.2 (sRGB) gamma is quite accurate for dark luminance regions, while at high luminance the logarithmic behavior prevails
HIGH DYNAMIC RANGE CAPTURE / ACQUISITION
HDR Video Capture / Acquisition

- **Single Sensor**
  - Ideal solution but expensive
  - 14 – 16.5 f-stops at 4K capture
  - Sensors may suffer from noise in low light

- **Temporal Bracketing**
  - Alternating frames 2-6 f-stops apart
  - Introduces ghosting artefacts in moving scenes
  - Typically requires high frame rate for video

- **Spatial Bracketing**
  - Beam splitters or mirrors redirect some of the light to additional sensors
  - Need to correct the geometric disparity between images
  - No temporal mismatches = no ghosting

- **Synthesis from Conventional Content**
Single Sensor

- **Every pixel captures a higher dynamic range**
  - Technically challenging
  - 14 – 16.5 f-stops of dynamic range claimed
  - Difficult to manage noise

- **Trade-off:**
  - Resolution
  - Noise
  - Dynamic Range

Sony F65

Arri Alexa
Single Sensor

- Light sensitivity of sensors can be increased by adding a white pixel

Single Sensor

- Each pixel captures different exposure
- Filters placed on the sensor elements

Single Sensor

- Bayer Strips (proposed for mobile sensors)
Single Sensor

‘Assorted pixels’

The arrangement of pixels can be varied to achieve different effects:

- **Dynamic range**
- **Wider gamut**
- **Polarization**
- **Etc…**

Single Sensor

- Adaptive Dynamic Range Imaging

Single Sensor

- **Control CMOS readout timing and exposure per row**
  - Use rolling shutter to our advantage
  - Well suited for images where exposure varies horizontally

Temporal Exposure Bracketing

- Each image has some well-exposed pixels
- **Merge into HDR:**
  - Apply inverse camera response function $f()$ to pixel value $Z$
  - Divide by exposure time $t$
  - Mask under- and over-exposed pixels with weight function $w$
  - Sum exposures

$$L(x, y) = \frac{\sum_{i=1}^{n} w(Z_i(x, y)) \left( \frac{f^{-1}(Z_i(x, y))}{t_i} \right)}{\sum_{i=1}^{n} w(Z_i(x, y))}$$

- **Method is effective if:**
  - Camera is motionless
  - Nothing moves in the scene
- **Hence: much research devoted to removing these limitations!**
Camera Response Recovery

- Cameras have non linear response
- To merge exposures, they need to be linearized first
- Requires recovery of the camera response curve $f()$

- Several approaches:
  - Mann & Picard (1995)
  - Debevec & Malik (1997)
  - Mitsunaga & Nayar (1999)
  - Robertson et al. (2003)
Temporal Exposure Bracketing

- **Two or more exposures taken one after the other**
  - Exposure times varied by a fixed number of f-stops (exponential)
  - More optimal spacing: using Fibonacci sequence – facilitates image registration (Gupta, Iso and Nayar, ICCV 2013)
  - Metering-based spacing

- **Possible even on smartphones**
- **Not well suited to dynamic content**
- **Trade off:**
  - Ghosting vs dynamic range

- **E.g. RED Epic Cameras (HDRx)**
  - 18.5 f-stops claimed for RED Epic Mysterium
Temporal Exposure Bracketing

- **Multiple shots taken of the same scene**
- **Exponential sequence:**
  - Exposure time multiplies by a factor between each image
  - For example: factor of 2 (i.e. 1 f-stop)
- **Most modern cameras offer automated settings (AEB - Auto-exposure bracketing)**
Temporal Exposure Bracketing

- Each exposure captures part of the scene
- Exposure stack can be merged into HDR image

Visualisation:
- Bit-pattern for each channel:
  - 1: well exposed
  - 0: under or over exposed
Temporal Exposure Bracketing

- **Merge exposure sequences into HDR imagery using Patch Match**
  - Finds and merges patches from different exposures
  - Extends to video by combining optical flow techniques with patch-based reconstruction

- **Advantages**
  - Implicit exposure alignment
  - Implicit ghost removal


Joint Deghosting and Reconstruction

- Alternatively we can try reconstruct new exposures where features are aligned
  - Patch-wise alignment to reference image

Joint Deghosting and Reconstruction

- This can be extended to video
  - Each exposure captured leads to one HDR frame
  - No loss of framerate

Metering Strategies

- Typically exposures are evenly spaced
  - For example: -2, 0, 2 f-stops
Metering Strategies

- **Metering allows scene-adaptive exposure selection**
  - Uses histogram analysis
  - For noise reduction
  - For capturing a higher dynamic range with fewer exposures

Metering Strategies

• First capture many exposures fast at a lower resolution
• Then reconstruct a histogram over the full dynamic range of the scene
• Finally, determine exposures that minimize noise, maximize dynamic range etc.
• Shown to reduce the number of exposures for an equivalent dynamic range
• However, histogram construction takes time
  ▪ Not suitable for video

Solution for HDR video capture on mobile devices

Uses previously captured frames for (partial) metering
- Steps up and down through exposure times based on histogram and motion analysis

Algorithm outline:
- Consider 2 previous frames
- Estimate motion
- Define threshold of ‘allowed’ under/over exposed pixels
- Use motion and under/over exposure information to determine how to expose the next frame

Off-line processing: adapted form of patch-math

Result: HDR video with frame rate identical to capture rate

## Example results

<table>
<thead>
<tr>
<th>[-3, 0]</th>
<th>[-2,0,2]</th>
<th>With metering</th>
<th>Ground truth</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image 1" /></td>
<td><img src="image2.png" alt="Image 2" /></td>
<td><img src="image3.png" alt="Image 3" /></td>
<td><img src="image4.png" alt="Image 4" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Image 5" /></td>
<td><img src="image6.png" alt="Image 6" /></td>
<td><img src="image7.png" alt="Image 7" /></td>
<td><img src="image8.png" alt="Image 8" /></td>
</tr>
<tr>
<td><img src="image9.png" alt="Image 9" /></td>
<td><img src="image10.png" alt="Image 10" /></td>
<td><img src="image11.png" alt="Image 11" /></td>
<td><img src="image12.png" alt="Image 12" /></td>
</tr>
</tbody>
</table>
Potential Problems with Temporal Bracketing

- **Noise**
  - Division by exposure time may amplify noise

- **Image alignment**
  - Hand-held exposures may be misaligned

- **Ghosting**
  - Objects may move between exposures
  - When reconstructing the HDR image, some information will be missing between exposures
  - Leads to ghosting artefacts
Spatial Bracketing

- Light can be split and redirected to different sensors
- Split may be done in front of the lens or behind the lens
  - Use multiple cameras, placed in front of a beam splitter
  - Split light after the main lens in a single camera containing a beam splitter
- Exposures are captured simultaneously
  - Optical filters control the amount of light reaching each sensor
- Advantages
  - No temporal mismatches
  - No ghosting
Example: NEVEX HDR Capture

- Based on modified stereo capture rig
- Beam splitter redirects light to second camera
- Cameras must be synchronized and aligned
- Disadvantage: cumbersome post-processing
Example: Contrast HDR Camera

- Capture multiple LDR images
  - Beam splitter redirects light to various sensors
  - Beam splitter attenuates light by different amounts dependent on direction
- Structurally identical images
- In theory any sensor could be used
- Large dynamic range
- Single camera lens
- Light-efficient

Summary

- **Multi-camera approaches impractical**
  - Multiple cameras with lenses means high material costs
  - Alignment / registration means significant (possibly manual) postprocessing
- **Multi-exposure approaches artifact-prone**
  - Ghost removal will be necessary in practical situations
- **Spatial bracketing provides good trade-off**
  - Multiple sensors are factory-aligned
  - All technology hidden from the user: operation should be no more straightforward
- **Synthesis from SDR content**
  - Most likely professionally employed solution for the foreseeable future (episodic content, movies, live broadcast)
  - In essence, this is the inverse of tone reproduction
  - Discussed later in this tutorial
TONЕ REPRODUCTION
Tone Reproduction

- Consider an HDR image that has a range of values that is much larger than a given display can display
  - Some range reduction will need to be applied
  - On some sense this means loss of information must occur

- The main research question is which information must be maintained, and which information is visually less pertinent and could be removed
  - Preserve details?
  - Preserve contrast?
  - Preserve visual appearance?
Why do we need tonemapping?

Linear compression

HDR luminance color map
(4 orders of magnitude - 13 f-stops)

Tonemapped
Linear Scaling

The diagram illustrates the relationship between input luminance and tone-mapped luminance. The graph shows a linear scaling where the tone-mapped luminance is directly proportional to the input luminance.
Linear scaling
Linear + Clamping

Tone Mapped Luminance vs. Input Luminance graph.
Linear + Clamping
$L_d = \log(L + 1)$
Logarithm
Power Function (gamma)

- Smaller power gives more compression

\[ L_d = L^\gamma \]
Power Function – gamma = 0.75
Power Function – gamma = 0.50
Power Function – \( \text{gamma} = 0.25 \)
Perception-based TMOs

- Human visual system faces similar problem
- Photoreceptors compress light non-linearly
\[ L_d = \frac{L}{L + \sigma} \]
$L_d = \frac{L}{L + \sigma}$
Sigmoid - $\sigma = 10$
Sigmoid - $\sigma = 1$
Sigmoid - $\sigma = 0.1$
Perception-based TMOs

- Sigmoidal (Reinhard 2002, 2005)
- Display-adaptive (Mantiuk 2008)
- Full photoreceptor model (Pattanaik 2000)
Global operators

- Recap:
  - Easy to implement!
  - Real time implementations possible
  - Some provide very reasonable amounts of compression and lead to plausible results
  - Useful for medium to high dynamic range images
  - No local compression means flat appearance
Global to local

- Global
Global to local

• Local
Simple solution:
Replace global parameter (e.g. $\sigma$) with local average

For sigmoid: $\sigma$ based on local neighborhood determined through low pass filtering:

$$L_d(x, y) = \frac{L(x, y)}{L_{LPF}(x, y) + L(x, y)}$$

How do we decide the size of local region?
Local Tone Mapping through Low Pass Filter

- For each pixel the size of the LPF kernel should be such that it does not overlap with sharp discontinuities.
- But we still want to average over the largest spatial area for which the above is true (which may be different for each pixel):
  - Flat areas = large possible kernel
  - High contrast areas = smaller kernel

- Too large a kernel can cause halos!
- Too small a kernel reduces local contrast.
Halos
How to compute?

- Multi-scale analysis using difference of Gaussians to compute local kernel size (Reinhard 2002, Ashikhmin 2002)
- Multi-scale decomposition (bilateral filter, mean shift algorithm, weighted least squares filtering etc)
- Distance to light sources (Reinhard et al 2012)
Multi-layer Decompositions

- Instead of preserving local detail directly, it can be removed first and then added back after compression

- Two or more layer decompositions
  - Coarse levels encode global contrast
  - Finer levels encode local contrast and detail

- Different creative effects can be achieved by compressing each level differently
Base/Detail Decomposition

- Split image into base and detail layers
- Base layer: image luminance filtered using edge stopping filter
- Detail layer: residual
Base/Detail Decomposition
Gradient Domain Compression

- **Compute image gradients in log space**
- **Large gradients = large contrasts**
  - By attenuating only large gradients, high contrasts are compressed but local contrast is preserved
- **Reconstruct image by integrating gradients**
  - By numerically solving a Poisson equation

Gradient Domain Compression
Informal Comparison

- Linear scaling
- Linear with clamping
- Linear with clamping (pct)
Informal Comparison

Reinhard 2002  Ashikhmin 2002  Li 2005
Informal Comparison

Histogram adjustment  Photoreceptor-based  Tumblin-Rushmeier
Inverse Tone Mapping

Goal:
Increase the dynamic range of content to match that of an HDR display, to improve visual the experience while preserving artistic intent.
Inverse Tone Mapping

- **Conceptually the reverse problem of tone mapping:**
  - Begin with a standard dynamic range (SDR) image
  - Recreate a high dynamic range image from it

- **Possible approach**
  - Many tone mappers use monotonically increasing curves
  - Take any such tone mapping operator, and invert it

- **Things to look out for**
  - Management of director’s intent
  - Blacks should not be pulled apart
  - Mid-tones should not be raised too much
  - Artefacts should not be amplified, esp. noise
Inverse Tone Mapping

- What is the best way to expand the range?
- Several psychophysical studies conducted
  - Banterle et al.: Inverse sigmoid
  - Masia et al.: Image dependent gamma
  - Akyuz et al.: Linear

  - Global vs Local debate, like for tone mapping

- Technicolor:

- Opto-Electrical Transfer Functions (OETF)
  - SMPTE 2084
  - ITU-R BT.2100
Global Inverse Tonemapping

- Single curve to expand luminance
  
  \[ L_d(x, y) = f(L(x, y)) \]

- All pixels treated the same
- Examples:
  
  - Solve: \( \left( \frac{\alpha}{L_w} (1 - L_d(x, y)) \right)^2 + \frac{4\alpha^2}{\beta L_w} L_d(x, y) \geq 0 \)

  - Compute: \( L_d(x, y) = L(x, y)^{g(L(x,y))} \)

- Computationally efficient, but perhaps not enough control in practice
Local Inverse Tonemapping

- Detect highlight regions
- Global expansion to all pixels
- Extra expansion for highlights

**Challenge:**
- Distinguish specularities, light sources and diffuse white areas
Highlight Expansion

- Most methods dedicate most of the range to highlights
- Compute brightness map to determine location of highlights
  - Median cut (Banterle 2006)
  - Multi-scale analysis (Rempel 2007)
- Modulate expansion depending on brightness map
Inverse Tonemapping – Additional Considerations

- **How to deal with noise?**
  - Expanding content may also expand noise
  - Artefacts that were below visible threshold, may become visible

- **How to deal with compression artefacts?**
  - Compression algorithms are very efficient at keeping artefacts just below visible threshold
  - Even slight expansion may render such artefacts visible

- **How to deal with over-exposed areas?**
  - Over-exposed areas are by definition featureless
  - Expanding such areas will make them very conspicuous and distracting
  - ‘Declipping’ may be applied with some success
Summary

- **Tonemapping**
  - Much research has led to good, practical solutions

- **Inverse Tonemapping**
  - Research in this area remains sparse
  - Basic expansion functions are relatively straightforward to design
  - Dealing with amplification of artefacts remains a challenge
COLOR MANAGEMENT
Color Management for HDR

- Color appearance changes if we only modify luminance!
Schlick Correction

- Exponent $s$ controls saturation
- Can cause luminance & hue shifts

$$C_{out} = \left( \frac{C_{in}}{L_{in}} \right)^s L_{out}$$

Schlick’s Method

Saturation parameter $s$

lower  higher
Mantiuk Correction

- Reduces luminance shifts
- May cause hue shifts
- Parameter $s$ can be determined based on the slope of the tone mapping curve

$$C_{out} = \left(\left(\frac{C_{in}}{L_{in}} - 1.0\right) s + 1.0\right)L_{out}$$

R. Mantiuk, R. Mantiuk, A. Tomaszewska and W. Heidrich, ‘Color Correction for Tone Mapping’, Eurographics 2009
Mantiuk’s Method

Saturation parameter $s$
Color Correction

- Both Schlick and Mantiuk corrections require manual parameter adjustment
- Restricted to global tonemapping operators

- Alternatively, given input and tone mapped image we can analyze saturation and try to preserve it:

\[ \text{Saturation} = \frac{\text{Chroma}}{\text{Lightness}} \]
Color Correction using Input/Tone mapped Pair

Comparisons

<table>
<thead>
<tr>
<th>Tonemapped input</th>
<th>Our method</th>
<th>Schlick</th>
<th>Mantiuk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Li 2005</td>
<td>Reinhard 2002</td>
<td>Linear HDR</td>
<td>Linear HDR</td>
</tr>
<tr>
<td>Li 2005</td>
<td>Reinhard 2002</td>
<td>Linear HDR</td>
<td>Linear HDR</td>
</tr>
<tr>
<td>Li 2005</td>
<td>Reinhard 2002</td>
<td>Linear HDR</td>
<td>Linear HDR</td>
</tr>
</tbody>
</table>
HDR VIDEO CODING
### HDR formats

- **HDR images**
  - Physical values of luminance
  - Represented using floating point
  - Cover full color gamut
  - Significantly higher raw data rate
  - Not appropriate for storage or transmission

<table>
<thead>
<tr>
<th>HDR Formats</th>
<th>bpp</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenEXR (.exr)</td>
<td>48 (half-float), 96 (float)</td>
</tr>
<tr>
<td>RGBE, XYZE (.hdr)</td>
<td>32 (float)</td>
</tr>
<tr>
<td>LogLuv TIFF (.tiff)</td>
<td>24, 32 (integer)</td>
</tr>
</tbody>
</table>
Compression: general concepts

- Exploit the correlation in the data
  - Reduce redundancies
  - Lossless

- Exploit the human visual system
  - Remove imperceptible data
  - Introduce (hardly) noticeable distortions
  - HDR: adapt human visual system models

- Performance measured in terms of rate-distortion
Timeline of video coding standards

- **MPEG**: Moving Picture Expert Group
  - ISO: International Standardization Organization
  - IEC: International Electrotechnical Commission

- **VCEG**: Video Coding Expert Group
  - ITU: International Telecommunication Union
Scope of video coding standards

- **Minimum for interoperability**
  - Syntax and semantic of the compressed bitstream
  - Decoding process
  - Leaves room for competition in the market place and future improvements
HEVC (MPEG-H – H.265)

- **Hybrid video coding**
  - Transform coding: DCT-like transform to compact the energy of the signal
  - Predictive coding: Intra or Inter (motion compensated) prediction
  - Entropy coding: context-adaptive binary arithmetic coding (CABAC)
**HEVC**

- **Quad-tree decomposition**
  - Coding Tree Units (CTU)
  - Coding Units (CU)
  - Prediction Units (PU)
  - Transform Units (TU)
HEVC

• **First version completed in 2013**
  – Approx. 50% bit rate savings compared to its predecessor H.264/AVC at the same quality
  – Main profile: 8 bit depth, 4:2:0 chroma sampling
  – Main 10 profile: 10 bit depth, 4:2:0 chroma sampling -> UHDTV

• **Second version in 2014**

• **Range extensions (RExt)**
  – More than 10 bits per sample
  – 4:0:0, 4:2:2 and 4:4:4 chroma sampling

• **Scalability extensions of HEVC (SHVC)**
  – Encode once, decode at different resolution/quality levels
  – Backward-compatible base layer, enhancement layers
  – Interlayer prediction mechanisms
  – Bit depth and color gamut scalability, also handles different transfer function
Wide Color Gamut - BT.2020 vs BT.709

- **BT.709** specifies formats for HDTV
- **BT.2020** specifies formats for UHDTV
  - Resolution, frame rate, chroma subsampling, bit depth, and color space
  - **BT.2020** covers a wider color gamut

CIE 1931 chromaticity diagram
OETF & EOTF

- **Opto-Electronic Transfer Function (OETF)**
  - Non-linear transfer function implemented within the camera

- **Electro-Optical Transfer Function (EOTF)**
  - Non-linear transfer function implemented within the display
BT.1886 – Gamma encoding

- **EOTF for television**
  - Well suited for a peak level of 100 nits, i.e. below visual detection thresholds
  - 12 bit gamma curve is above Schreiber and Barten thresholds for higher peak luminance levels

**Schreiber** threshold: Weber’s law above 1 cd/m² and gamma nonlinearity below

**Barten** threshold: based on model of Contrast Sensitivity Function

S. Miller, M. Nezamabadi, S. Daly, Perceptual Signal Coding for More Efficient Usage of Bit Codes
2012 SMPTE Annual Technical Conference & Exhibition
Perceptual Quantizer (PQ) - SMPTE ST 2084

- Derive an EOTF based on perception
  - Maximize the dynamic range of the signal by setting each quantization step to be proportional to the Just Noticeable Difference (JND)
  - Iterative EOTF computation, exploiting the Barten CSF model
  - In the range 0 to 10000 nits

Can be approximated by the model, known as PQ (SMPTE ST 2084):

$$Y = L \left( \frac{V^{1/m} - c_1}{c_2 - c_3 V^{1/m}} \right)^{1/n}$$

12-bit PQ remains below perceptual thresholds

S. Miller, M. Nezamabadi, S. Daly, Perceptual Signal Coding for More Efficient Usage of Bit Codes
2012 SMPTE Annual Technical Conference & Exhibition
Perceptual Quantizer (PQ) - SMPTE ST 2084

10 bit video is still the norm; HEVC Main 10 profile is foreseen for distribution

10-bit PQ still near or below perceptual thresholds

S. Miller, M. Nezamabadi, S. Daly, Perceptual Signal Coding for More Efficient Usage of Bit Codes
2012 SMPTE Annual Technical Conference & Exhibition
**Hybrid Log-Gamma (HLG) - ARIB STD-B67**

- Designed for backward compatibility with the BT.709 OETF and the corresponding BT.1886 EOTF with gamma encoding
  - The same content can be displayed on both LDR and HDR displays (at least in principle)
  - Working well for peak luminance up to approx. 1000 cd/m²

Inflexion point at relative luminance value $\mu$
- Below $\mu$, the curve is similar to the BT.709 transfer function
- Above $\mu$, the curve is stretched

Both PQ and HLG are currently standardized in ITU-R BT.2100

---

E. François, P. Bordes, F. Le Léannec, S. Lasserre, P. Andrivon
High Dynamic Range and Wide Color Gamut Video Standardization — Status and Perspectives
Heterogeneous Environment

Distribution

HDR Content
- 10 / 12 bit
- PQ / HLG
- Wide Color Gamut
  - Up to 10000 cd/m²

HDR Display
- Wide Color Gamut
  - Up to 10000 cd/m²

UHD TV
- BT.2020
  - Up to 1000 cd/m²

HD TV
- BT.709
  - Up to 100 cd/m²
Different distribution scenarios

**Single-layer**
- HDR video coding using HEVC Main 10
- HDR-to-LDR mapping at the receiver side (real-time)
Different distribution scenarios

- **Single-layer**
  - HDR video coding using HEVC Main 10
  - HDR-to-LDR mapping at the receiver side (real-time)
  - Metadata from LDR master video to improve tone/color mapping and preserve artistic intent
Different distribution scenarios

- **Single-layer**
  - LDR video coding using HEVC
  - LDR-to-HDR mapping at the receiver side (real-time)
Different distribution scenarios

- **Single-layer**
  - LDR video coding using HEVC
  - LDR-to-HDR mapping at the receiver side (real-time)
  - Metadata from HDR master video to improve inverse tone/color mapping and preserve artistic intent

![Diagram of HDR distribution scenarios](image)
Different distribution scenarios

- **Single-layer**
  - LDR video coding using HEVC
  - LDR-to-HDR mapping at the receiver side (real-time)
  - LDR video obtained by HDR-to-LDR tone/color mapping
Different distribution scenarios

- **Dual-layer, simulcast**
  - Simulcast (independent coding) of LDR and HDR video sequences
  - Not optimal in terms of bandwidth
Different distribution scenarios

- **Dual-layer, scalable coding**
  - Joint coding of LDR and HDR video sequences using SHVC
  - HDR enhancement layer can be predicted from LDR base layer
  - Exploit better data redundancies
MPEG approach

Input HDR video
Linear R, G, B → Coding TF → R’G’B’ To Y’CbCr → Quant. 10b → 4:4:4 To 4:2:0 → Encoding HEVC Main 10 → Transmission

Output HDR video
Linear R, G, B → Inv. Coding TF → Y’CbCr To R’G’B’ → Inv. Quant. 10b → 4:2:0 To 4:4:4 → Decoding HEVC Main 10
MPEG approach

- **SMPTE 2084/PQ**
  - Comparison of 8-, 10- and 12-bit HEVC coding
Dual-modulation scheme

- Two multiplicative layers
- Assume that the dynamic of the HDR image is limited within a small spatial area
- The modulation signal is smooth and can be downsampled
- Mimic the process of HDR displays

$$P_{\text{HDR}} = P_{\text{mod}} \times f^{-1}(P_{\text{LDR}})$$

*P_{\text{mod}}*: low-frequency monochromatic version of the input signal

*P_{\text{LDR}}*: remaining part of the signal

---

E. François, P. Bordes, F. Le Léannec, S. Lasserre, P. Andrivon
High Dynamic Range and Wide Color Gamut Video Standardization — Status and Perspectives
Dual-modulation scheme

\[ P_{\text{HDR}} = P_{\text{mod}} \times f^{-1}(P_{\text{LDR}}) \]
Schemes based on content-adaptive TMO

- **Content-adaptive TMO**
  - Take into account the statistical characteristics of the input HDR frame
  - Metadata need to be transmitted, but with a negligible overhead cost

![Diagram showing the process of TMO encoding and iTMO decoding with metadata.]
### MSE optimal TMO

- **Piecewise linear TMO**

- **TMO minimizing the MSE between original and reconstructed log values**

\[
\min \sum_{k=1}^{N} p_k s_k^{2-2} \quad \text{s.t.} \quad \sum_{k=1}^{N} s_k = \frac{v_{\max}}{\delta}
\]

- **Closed form solution**

\[
s_k = \frac{v_{\max} \cdot p_k^{1/3}}{\delta \cdot \sum_{k=1}^{N} p_k^{1/3}}
\]

where \( p_k \) is the \( k \)th bin of the normalized luminance histogram.

---

Spatial regularization

- **Optimization problem**

\[
\min_s \left( \sum_{k=1}^{N} p_k s_k^{-2} + \lambda \text{TV}(I(s)) \right) \quad \text{s.t.} \quad \sum_{k=1}^{N} s_k = \frac{v_{\text{max}}}{\delta}
\]

- **Spatial regularization term**
  - Total variation

\[
\text{TV}(I(s)) = \sum_i \| (\nabla I)_i \|
\]

- **Non-smooth convex function**
  - No closed-form solution
  - Primal-dual M+LFBF algorithm [Combettes and Pesquet, 2012]

Temporal regularization

- Optimization problem

\[
\min_s \left( \sum_{k=1}^{N} p_k s_k^{-2} + \lambda C(I_t) \right) \quad \text{s.t.} \quad \sum_{k=1}^{N} s_k = \frac{v_{\text{max}}}{\delta}
\]

- Temporal regularization term

\[
C(I_t) = \sum_{i,j} (I_t(i, j) - M(I_{t-1}(i, j)))^2
\]

- \( M(I_{t-1}(i,j)) \) is the motion compensated frame
  - Optical flow computed on the HDR frames at time \( t \) and \( t-1 \)

Frame difference

- MSE optimization only
- With temporal regularization

Carousel_Fireworks_03
Rate-distortion results using HDR-VDP 2.2.1

[Tutorial on HDR Video, EUSIPCO 2016]
Rate-distortion results using HDR-VDP 2.2.1

![Graphs showing rate-distortion results using HDR-VDP 2.2.1.](Carousel_Fireworks_03, Carousel_Fireworks_04)
JPEG family of standards

- **JPEG 2000**
  - State-of-the-art still image coding scheme, based on a discrete wavelet transform (DWT)
  - Bit-depth in the range of 1 to 38 bits per component

- **JPEG XR**
  - Cost-effective compression, based on block-transform design
  - Different image representation pixel formats: 8- and 16-bit unsigned integer, 16- and 32-bit fixed point, 16- and 32-bit floating-point

- **JPEG XT**
  - Backward compatible extensions to the legacy JPEG standard for HDR image compression
  - Two layers design
  - Base layer: tone-mapped and encoded using JPEG
  - Extension layer: residual information by subtraction and/or division
JPEG family of standards

- JPEG XT Profile C
  - Residual = the ratio of the HDR image and the inverse tone-mapped image, implemented as a difference of logarithms
HDR DISPLAY TECHNOLOGY
Motivation

- Displaying HDR images and videos directly, without tone mapping

- Conventional Liquid Crystals Displays (LCD) technology
  - Max luminance ~ 300-400 cd/m2
  - Max dynamic range ~ 500:1
  - Human eye can perceive a much higher range of luminance, even at a single state of adaptation

- Questions:
  - What is the needed peak luminance?
  - What is the needed dynamic range?
  - How to build hardware solutions to achieve them?
Peak brightness alone is not sufficient

- Viewers prefer higher peak brightness *provided that* dynamic range increases concurrently (Akyuz, 2006; Seetzen, 2006)
- Problem of the *black level*
  - If the display dynamic range (also called contrast ratio) is fixed, increasing peak brightness rises the black level

[Seetzen et al., 2006]
What is the required dynamic range?

- **Lower bound: SSDR (steady-state dynamic range)**
  - About 4 log$_{10}$ units (~ 13 f stops)
  - Can be larger when looking at full images (changes with eye movements, viewing angle, etc.)

- **Upper bound: long-term adaptation**
  - Up to 14 log$_{10}$ units
  - Unrealistic to reproduce

- **The necessary dynamic range is somewhat in-between**
What is the required dynamic range? (cont’d)

[Diagram showing luminance levels and viewer preference distribution]

- Black stimuli: Current cinema — ~12 f-stops
- White stimuli: Standard TV — ~10 f-stops
- Highlights: Large screen — viewer preference dynamic range — 22 f-stops

Legend:
- Small screen study
- Large screen study
- Inferred data
- Outside large screen DR
- Outside small screen DR

[Kunkel et al., 2016]
Local contrast perception

- **How to reproduce 5 or more orders of magnitude of dynamic range simultaneously?**
  - No conventional display technology is able to achieve this contrast

- **Fortunately, it is not necessary!**
  - The actually visible dynamic range in small regions (small visual angles) is much smaller

- **Veiling glare effect:**
  - The point spread function of the eye is not an impulse
  - A fraction of the light is scattered on neighboring receptors, lowering overall contrast
  - Maximum perceivable contrast ~ 150:1 (Seetzen et al., 2004)

[Point spread function of the eye][Seetzen et al., 2004]
Hardware solutions to display HDR

- **Basic idea: two-layer solution** (*dual modulation*)
  - The first layer (*light modulator*) emits high-intensity, but low-resolution, light
  - The second layer (*optical filter*) filters the light with high-frequency details

- **If** $c_1$ **is the contrast ratio of the first layer,** $c_2$ **that of the second layer**
  - The theoretical maximum dynamic range achievable is $(c_1 \cdot c_2) : 1$
  - Limited by physical transmittance of the second layer

- **The optical filter is generally a LCD screen**
- **For the light modulator, two solutions** (*Seetzen et al., 2004*):
  - Digital Light Projector (DLP)
  - Array of ultra-bright LED’s
Projector-based display

• Composed of a projector, an LCD display and a diffuser
  – The latter to avoid Moiré patterns
  – Achievable dynamic range > 50,000:1

• Very high image quality
  – Relative high-resolution of projector image
  – Adapt for research applications as a home-made solution

• Not practical for the consumer market
  – Form factor
  – High energy consumption
  – Projector cost
  – Higher bandwidth

[Seetzen et al., 2004]
LED-based display

- **Composed by the LED layer, diffuser and LCD panel**
  - LED’s individually controllable, positioned in an hexagonal pattern
  - Diffuser avoids discontinuities in LCD illumination

- **W.r.t. projector-based:**
  - Reduced form factor
  - Higher power efficiency
  - Lower bandwidth (limited number of LED’s, 2200 for SIM2 display)
  - Higher computational complexity in the rendering due to the much larger support of the point spread function (PSF) of the diffuser
Dual modulation rendering

- The front and back panels have to be driven with separate signals, computed through a *dual modulation* algorithm
  - The LCD signal compensates for the low-resolution backlight
- E.g.: Brightside rendering for LED-based display

[Diagram showing the process of dual modulation rendering]

[Trentacoste et al., 2007]
Target backlight computation

- Given the desired image $\hat{I}$, determine the target backlight distribution $\hat{B}$
  1. Compute image luminance as the maximum of the three color channels (to provide enough illumination for LCD)
  2. Clamp luminance values to be in the display luminance range
Target backlight computation

- **Given the desired image** $\hat{I}$, determine the target backlight distribution $\hat{B}$

1. Compute image luminance as the maximum of the three color channels (to provide enough illumination for LCD)
2. Clamp luminance values to be in the display luminance range
3. Low-pass filter the luminance of the image and (possibly) down-sample it on the LED grid to reduce computational complexity
Derivation of LED values

- Find the LED activation values such that the resulting backlight is as close as possible to $\hat{B}$
  - Essentially, a *deconvolution* problem
  - LED driving values $d \in [0,1]^N$, where, e.g., $N = 2202$ for SIM2 HDR47 display
  - Finding $d$ corresponds to solving the following (over-constrained) least-square problem:
    $$d^* = \arg \min_d \| Wd - \hat{B} \|$$
    where $W$ is a matrix containing the PSF of LED at each LED position
  - *Most complex operation of the rendering pipeline*

- Several possible approximated solutions:
  - Leverage the sparsity of $W$
  - Downsampling of $\hat{B}$ and doing one step of Gauss-Seidel iteration (Trentacoste et al., 2007)
  - Iterative scaling (Zerman et al., 2015)
Derivation of LED values

- Estimated driving values
- Clipped in [0,1] and quantized on 8 bits
- Power absorption constraint:

\[ \sum d_i P_{LED} \leq P_{MAX} \]

- The total absorbed power cannot exceed the maximum available power
- Re-scaling of LED values, or inclusion of the constraint in the optimization (Burini et al., 2013)
Forward backlight simulation

- Once the optimal LED driving values are known, the resulting backlight can be found by convolving the vector of Dirac’s $d^*$ with the PSF of LED’s:

$$\tilde{B} = d^* \ast PSF$$

- Can be efficiently implemented in GPU through a splatting approach + alpha blending
Computation of LCD values

- Obtained by dividing the input target image by the computed backlight:
  \[ \text{LCD} = \frac{\hat{I}}{\hat{B}} \]
  - Performed per pixel and per color channel
  - The resulting LCD pixels are processed with the inverse response function of LCD panel (e.g., gamma), and clipped in [0, 255]

- Resulting LCD images have reversed contrast at edges (e.g., halo effects) to compensate for the backlight blur
Extension to video

- **Frame-by-frame rendering**
  - Changes in LED driving signals can produce flickering
  - The refresh time for LED and LCD panels are generally different, thus video rendering could be from unsynchronized frames

- **Spatial smoothing of backlight**
  - Fastest and most popular solution
  - Reduces the overall dynamic range

- **Temporal optimization algorithm**
  - Smooth backlight *temporally*
  - E.g., compute the envelope of luminance across backlight pixel trajectories (Zerman et al., 2016)
Guidelines for HDR content rendering

- **HDR content adaptation**
  - Although with a much wider dynamic range than LDR, HDR displays still have physical limitations
  - HDR content should be adapted (e.g., tone-mapped) to meet the display capabilities!
  - Clipping of specular highlights, manual grading, automatic display adaptive tone mapping (Mantiuk et al., 2008)

- **Modeling of the display**
  - Accurate measurement of the LED PSF
  - Characterization of power constraints, LCD gamma’s, location of LED grid, alignment issues for projector-based models, etc.

- **Display limitations:**
  - Trade-off between accuracy of reproduction, computational complexity, video flickering
  - Current solutions are mainly heuristic
  - Loss of contrast due to LCD leakage, lowered thresholds due to local adaptation, etc.
Laboratory HDR displays

- **SIM2 HDR47 series**
  - Peak luminance 4,000 or 6,000 nits (LED-based)
  - Dynamic range > 100,000 : 1
  - Full HD resolution
  - 60 Hz frame rate real-time rendering

- **High power consumption**
  - 2202 LED’s
  - 1380 W @ 4,000 nits
  - Full peak luminance available only on limited areas of the image to meet power constraints

- **Driving mode**
  - HDR mode: LogLUV with 12 bits for luminance
  - DVI+ mode: custom rendering
Consumer HDR displays

- Professional displays
  - E.g. Sony BVMX300
  - Used for professional HDR color grading

- Consumer displays
  - LED vs. OLED

- UHD alliance “premium 4K experience” logo
  - 1,000nits peak brightness and < 0.05nits black level (contrast ratio 20,000:1, ≈14.3 f-stops) for LED TVs which are brighter but with less black levels
  - > 540nits brightness and < 0.0005nits black level (contrast ratio 1,080,000:1, ≈20 f-stops) for OLED TVs which have deep blacks but much lower peak brightness.
HDR QUALITY OF EXPERIENCE
Quality of Experience (QoE)

- “The degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user’s personality and current state.” (Le Callet et al., Qualinet white paper, 2013)

- **Multidimensionality**
  - Immersiveness
  - Perceptual fidelity
  - Visual attention
  - Aesthetics and artistic intention
  - Naturalness, etc.

- **Measurement**
  - Subjective experiments
  - Computational models
Measurement of HDR QoE through subjective tests

- **Some important differences w.r.t. LDR**
- **Effect of display**
  - Effects of high luminance and contrast
  - Higher visual fatigue and discomfort
  - Grading of the content: clipping of highlights, tone mapping for display
- **Ambient luminance**
  - Recommendation ITU-R BT.500-13 recommends the illumination should be approximately 15% of display peak luminance
  - About 600 cd/m² for a 4,000 nits display!
  - Too high ambient illumination increases perceived black level causing loss of details in the dark parts of the image (Mantiuk et al., 2010)
  - Too low ambient illumination induces discomfort and fatigue
  - Reflections on screen surface increase with luminance as well
- **Practical recommendations for ambient luminance:**
  - 150-200 cd/m² (Narwaria et al., 2014)
  - 10-20 cd/m² (Valenzise et al., 2014; De Simone et al., 2014; MPEG tests)
HDR image and video quality prediction

- **Visual quality is one of the component of QoE**
  - Especially important in the HDR video delivery chain
  - Measure the performance of video compression
  - Fidelity of tone mapping when HDR has to be shown on a LDR display

- **Full-reference vs. No-reference quality metrics**
  - Perceptual fidelity vs. aesthetic judgements

- **Challenges w.r.t. LDR quality prediction**
  - HDR pixels encode scene luminance
  - Display rendering must be taken into account
  - The higher luminance renders noise more visible (contrast sensitivity is higher at higher)
Approaches for assessing HDR fidelity

- Error visibility metrics (near-threshold)
- Photometric values (HDR linear values)
- MOS predictors (supra-threshold)
- Perceptually uniform values
HDR-VDP 2 (Visible Difference Predictor, Mantiuk’11)

- Accurate simulation of the early stages of HVS, including an new CSF
- Designed and calibrated to detect visibility of errors, it can be applied as a suprathreshold quality metric (MOS predictor)
- Generalizes to a broad range of viewing conditions (scotopic/photopic)

[Mantiuk et al., 2011]
HDR-VDP 2 – Output

- **Visibility map** ($P_{map}$)
  - Per pixel probability of detecting a difference

- **Global quality score** ($Q_{MOS}$)
  - Value between 0 and 100
  - Obtained by pooling threshold-normalized subband differences
  - Pooling function obtained through model selection on two LDR training datasets
  - Version 2.2 of the metric has updated weights for the model computed on HDR content (Narwaria et al., 2015)

Reference Image | Test image | Probability of detection (screen, color)

[Mantiuk et al., 2011]
LDR metrics on perceptually encoded pixels

- Many quality metrics used on LDR images compute operations or functions of pixel values
  - Hypothesis: pixels can be compared (e.g., differences can be computed) because they lie in a perceptually uniform space
  - Conventionally, LDR images are encoded with a gamma (BT.1886 EOTF, sRGB) that simulates the nonlinear response of HVS

- No longer true for HDR pixel values
  - Pixel values are proportional to physical luminance
  - Luminance masking and adaptation should be taken into account

- To be compared in a perceptually meaningful way, HDR pixels have to be converted to perceptual units
  - Luminance to brightness transformation
    - Logarithm (Weber-Fechner law)
    - Power function (CIE LAB)
  - Perceptually uniform (PU) encoding (Ayding et al., 2008)
  - CSF-based transfer functions (e.g., PQ)
Extending LDR quality metrics to HDR

- Transform test and reference image pixels to perceptually uniform values before computing the quality metric
  - Differently from HDR-VDP, it is a simple luminance mapping
  - Computationally effective (look-up table)
  - The price to pay w.r.t. LDR is that the input of the mapping should be physical luminance in cd/m², i.e., a model of display is required

[Aydin et al., 2008]
Extensions to video – HDR-VQM (Narwaria’15)

- Suprathreshold video quality metric based on perceptually uniform encoded HDR pixel values
- Compute errors in frequency subbands, similar to HDR-VDP
  - Much simpler modeling (no retinal image, no inter-band contrast, etc.)
  - Much faster
- Spatio-temporal pooling
HDR-VQM Spatio-Temporal pooling

- Video divided in non-overlapping short-term spatio-temporal tubes
  - Spatial support to cover foveation area (approx. 2° of visual angle)
  - Temporal support: 300-500 ms
  - Standard deviation for short-term temporal pooling

- Spatial and long-term temporal pooling performed by excluding largest values from short-term temporal pooling

[Diagram of spatio-temporal pooling process]

[Source: Narwaria et al., 2015]
Dynamic range-independent metrics

- **Dynamic Range Independent (DRI) metric** (Aydin et al., 2008)
  - Visibility predictor
  - Adapted to evaluate images of different luminance and dynamic range (TMO’s, iTMO’s, display rendering)

- 3 structural changes can affect quality:
  - Loss of visible contrast
  - Amplification of invisible contrast
  - Reversal of visible contrast

- **Thresholds values:**
  - Visibility > 95%
  - Invisibility < 50%
DRI metric: evaluation of TMO’s

- **Output: distortion map**
  - Green: loss of visible contrast (e.g., details)
  - Blue: amplification of invisible noise (e.g., noise)
  - Red: contrast reversal (e.g., halos)

- **Suitable for visual inspection and qualitative evaluation**
- **Does not produce a global quality score**

[Aydin et al., 2008]
Tone-Mapped Image Quality Index (Yeganeh and Wang, 2013)

- Predicts overall quality of a tone-mapped image w.r.t. an HDR reference
- **Structural similarity \( S \):**
  - Modified SSIM index without luminance component
  - Contrast component adapted to detect only cases in which invisible contrast becomes visible or vice versa (similar to DRI)
  - Visual model employing a CSF and a psychometric function
- **Naturalness \( N \):**
  - Similarity between the histogram of tone-mapped image and the “average” histogram of LDR images
- **TMQI metric is the combination of the two terms:**

\[
Q = aS^\alpha + (1 - a)N^\beta
\]

- Spearman rank-order correlation coefficient (monotonicity of predictions) \( \sim 0.8 \)
Measuring performance of fidelity metrics

- Mean opinion scores predicted by quality metrics can be compared to actual MOS values collected in subjective studies.

- Statistical evaluation of metrics predictions (ITU-T P.1401)
  - **Accuracy** – Pearson’s correlation coefficient (PCC), root mean square error (RMSE)
  - **Monotonicity** – Spearman’s rank-order correlation coefficient (SROCC)
  - **Consistency** – Outlier Ratio (OR)
Benchmark of HDR image/video quality metrics

- Several studies on benchmarking HDR quality metrics
  - Different datasets
  - Different kinds of distortion (compression algorithms, TMO’s)
  - Different experimental procedures
  - Different conclusions!

- Extensive performance evaluation on 690 HDR images
  - 5 subjectively annotated datasets with aligned scores (Pinson and Wolf, 2003)

<table>
<thead>
<tr>
<th>Database</th>
<th>Obs.</th>
<th>Meth.</th>
<th>Stim.</th>
<th>Compr.</th>
<th>TMO</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 [18]</td>
<td>27</td>
<td>ACR-HR</td>
<td>140</td>
<td>JPEG(^1)</td>
<td>iCAM06 [27]</td>
</tr>
<tr>
<td>#3 [32]</td>
<td>24</td>
<td>DSIS</td>
<td>240</td>
<td>JPEG-XT</td>
<td>RG [30] \n</td>
</tr>
<tr>
<td>#4 [16]</td>
<td>15</td>
<td>DSIS</td>
<td>50</td>
<td>JPEG(^1) \n</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>JPEG 2000(^1)</td>
<td>Mai [34]</td>
</tr>
<tr>
<td>#5 [5]</td>
<td>15</td>
<td>DSIS</td>
<td>50</td>
<td>JPEG(^1) \n</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mai [34]</td>
<td>PQ [21], [35]</td>
</tr>
</tbody>
</table>
Benchmark of HDR image/video quality metrics

<table>
<thead>
<tr>
<th>Metric</th>
<th>Database #1</th>
<th>Database #2</th>
<th>Database #3</th>
<th>Database #4&amp;5</th>
<th>Combined</th>
<th>Except Database #2</th>
<th>Database #1</th>
<th>Database #2</th>
<th>Database #3</th>
<th>Database #4&amp;5</th>
<th>Combined</th>
<th>Except Database #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Photometric-MSE</td>
<td>0.4051</td>
<td>0.1444</td>
<td>0.7080</td>
<td>0.5095</td>
<td>0.3742</td>
<td>0.6292</td>
<td>0.750</td>
<td>0.933</td>
<td>0.787</td>
<td>0.830</td>
<td>0.830</td>
<td>0.762</td>
</tr>
<tr>
<td>Photometric-PSNR</td>
<td>0.4409</td>
<td>0.2564</td>
<td>0.7132</td>
<td>0.5594</td>
<td>0.4967</td>
<td>0.6507</td>
<td>0.771</td>
<td>0.905</td>
<td>0.767</td>
<td>0.820</td>
<td>0.813</td>
<td>0.717</td>
</tr>
<tr>
<td>Photometric-SSIM</td>
<td>0.5016</td>
<td>0.3583</td>
<td>0.8655</td>
<td>0.6708</td>
<td>0.6220</td>
<td>0.7596</td>
<td>0.821</td>
<td>0.938</td>
<td>0.679</td>
<td>0.780</td>
<td>0.786</td>
<td>0.694</td>
</tr>
<tr>
<td>Photometric-IFC</td>
<td>0.7781</td>
<td>0.8234</td>
<td>0.9183</td>
<td>0.8195</td>
<td>0.8153</td>
<td>0.8132</td>
<td>0.750</td>
<td>0.871</td>
<td>0.546</td>
<td>0.610</td>
<td>0.661</td>
<td>0.596</td>
</tr>
<tr>
<td>Photometric-UQI</td>
<td>0.7718</td>
<td>0.8208</td>
<td>0.8846</td>
<td>0.7876</td>
<td>0.8102</td>
<td>0.8113</td>
<td>0.707</td>
<td>0.871</td>
<td>0.558</td>
<td>0.640</td>
<td>0.658</td>
<td>0.623</td>
</tr>
<tr>
<td>Photometric-VIF</td>
<td>0.7603</td>
<td>0.5076</td>
<td>0.8666</td>
<td>0.6144</td>
<td>0.6450</td>
<td>0.7916</td>
<td>0.679</td>
<td>0.948</td>
<td>0.617</td>
<td>0.800</td>
<td>0.787</td>
<td>0.652</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>Database #1</th>
<th>Database #2</th>
<th>Database #3</th>
<th>Database #4&amp;5</th>
<th>Combined</th>
<th>Except Database #2</th>
<th>Database #1</th>
<th>Database #2</th>
<th>Database #3</th>
<th>Database #4&amp;5</th>
<th>Combined</th>
<th>Except Database #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-MSE</td>
<td>0.6114</td>
<td>0.5314</td>
<td>0.8856</td>
<td>0.8820</td>
<td>0.6464</td>
<td>0.7953</td>
<td>0.843</td>
<td>0.924</td>
<td>0.592</td>
<td>0.570</td>
<td>0.693</td>
<td>0.588</td>
</tr>
<tr>
<td>Log-PSNR</td>
<td>0.6456</td>
<td>0.5624</td>
<td>0.8870</td>
<td>0.8819</td>
<td>0.6671</td>
<td>0.8011</td>
<td>0.786</td>
<td>0.919</td>
<td>0.588</td>
<td>0.580</td>
<td>0.751</td>
<td>0.598</td>
</tr>
<tr>
<td>Log-SSIM</td>
<td>0.8965</td>
<td>0.8035</td>
<td>0.9235</td>
<td>0.8255</td>
<td>0.8235</td>
<td>0.8656</td>
<td>0.643</td>
<td>0.876</td>
<td>0.525</td>
<td>0.570</td>
<td>0.671</td>
<td>0.535</td>
</tr>
<tr>
<td>Log-IFC</td>
<td>0.7919</td>
<td>0.8366</td>
<td>0.9167</td>
<td>0.8551</td>
<td>0.8322</td>
<td>0.8448</td>
<td>0.750</td>
<td>0.833</td>
<td>0.529</td>
<td>0.610</td>
<td>0.646</td>
<td>0.575</td>
</tr>
<tr>
<td>Log-UQI</td>
<td>0.7837</td>
<td>0.8268</td>
<td>0.8786</td>
<td>0.7830</td>
<td>0.8065</td>
<td>0.8052</td>
<td>0.671</td>
<td>0.843</td>
<td>0.579</td>
<td>0.630</td>
<td>0.662</td>
<td>0.613</td>
</tr>
<tr>
<td>Log-VIF</td>
<td>0.5079</td>
<td>0.6202</td>
<td>0.8354</td>
<td>0.7065</td>
<td>0.5919</td>
<td>0.7169</td>
<td>0.807</td>
<td>0.924</td>
<td>0.654</td>
<td>0.730</td>
<td>0.843</td>
<td>0.683</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric</th>
<th>Database #1</th>
<th>Database #2</th>
<th>Database #3</th>
<th>Database #4&amp;5</th>
<th>Combined</th>
<th>Except Database #2</th>
<th>Database #1</th>
<th>Database #2</th>
<th>Database #3</th>
<th>Database #4&amp;5</th>
<th>Combined</th>
<th>Except Database #2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDR-VDP-2.2 Q</td>
<td><strong>0.9048</strong></td>
<td><strong>0.5980</strong></td>
<td><strong>0.9499</strong></td>
<td><strong>0.9407</strong></td>
<td><strong>0.8163</strong></td>
<td><strong>0.9385</strong></td>
<td><strong>0.564</strong></td>
<td><strong>0.919</strong></td>
<td><strong>0.346</strong></td>
<td><strong>0.520</strong></td>
<td><strong>0.587</strong></td>
<td><strong>0.421</strong></td>
</tr>
<tr>
<td>HDR-VQM</td>
<td>0.8949</td>
<td>0.8059</td>
<td><strong>0.9589</strong></td>
<td><strong>0.9333</strong></td>
<td><strong>0.8871</strong></td>
<td><strong>0.9383</strong></td>
<td>0.514</td>
<td>0.895</td>
<td>0.383</td>
<td>0.530</td>
<td><strong>0.575</strong></td>
<td><strong>0.410</strong></td>
</tr>
<tr>
<td>mPSNR</td>
<td>0.6545</td>
<td>0.6564</td>
<td>0.8593</td>
<td>0.8587</td>
<td>0.7209</td>
<td>0.7974</td>
<td>0.771</td>
<td>0.895</td>
<td>0.667</td>
<td>0.610</td>
<td>0.706</td>
<td>0.633</td>
</tr>
<tr>
<td>tPSNR-YUV</td>
<td>0.5784</td>
<td>0.4524</td>
<td>0.8319</td>
<td>0.7788</td>
<td>0.6440</td>
<td>0.7675</td>
<td>0.800</td>
<td>0.952</td>
<td>0.625</td>
<td>0.670</td>
<td>0.752</td>
<td>0.646</td>
</tr>
<tr>
<td>CIE ΔE* 2000</td>
<td>0.6088</td>
<td>0.2553</td>
<td>0.7889</td>
<td>0.6082</td>
<td>0.4971</td>
<td>0.7550</td>
<td>0.743</td>
<td>0.924</td>
<td>0.675</td>
<td>0.760</td>
<td>0.828</td>
<td>0.688</td>
</tr>
</tbody>
</table>
Perspectives and challenges

- **Color**
  - All the metrics discussed above are color blind
  - Color difference metrics such as CIE $\Delta E_{2000}$ have been designed for LDR content
  - Do not take into account color appearance phenomena

- **Aesthetic quality**
  - Artistic intent (in TMO, content grading for display, etc.)
  - Automatic optimization of these tasks
  - Estimation of aesthetic attributes (Aydin et al., 2015)
  - In the case of HDR, predicting the perceived dynamic range and colorfulness (Hulusic et al., 2016)

- **Visual fatigue**
  - HDR viewing may increase visual fatigue and discomfort
  - Dependency on ambient illumination and display peak luminance (Rempel et al., 2009)
WRAP-UP
Perspectives and challenges

- **HDR imaging technology is quite mature**
  - Perceptual phenomena and their impact on Quality of Experience are well understood
  - Many tools for image acquisition and reproduction

- **Still many challenges to solve for video**
  - Efficient video coding, standardization, backward compatibility
  - Real-time high-quality display
  - Video tone mapping and inverse tone mapping
  - Color management
  - Quality of Experience and aesthetic evaluation, etc.

- **HDR is a fundamental requirement for true immersive entertainment and communication**
  - New and fascinating artistic choices in digital media creation
  - Increasing diffusion of HDR content
  - Big players in streaming, cinema and broadcasting actively involved in standardization
HDR beyond entertainment

- High potential for a number of applications
  - Computer vision
    - Capture of dark and bright details for video surveillance
    - Automated driving and driver assistance systems
  - Simulation
    - Realistic driving simulation
    - Virtual reality
  - Medical imaging
    - Importance of accurate reproduction of gray levels
    - Design of novel luminance mapping for dark details
HDR for computer vision: image matching

- Illumination changes constitute a major challenge in video surveillance and analysis applications
  - Image matching is seriously affected by drastic changes in illumination
  - HDR can help to increase keypoint repeatability and descriptor matching
  - TMO’s have to be optimized for this purpose

A. Rana, G. Valenzise, F. Dufaux, Evaluation of Feature Detection in HDR Based Imaging Under Changes in Illumination Conditions, ISM 2015
Safety in automated driving

- Florida, 30th of June 2016: first fatal Tesla car accident
- The autopilot sensors on the Model S failed to distinguish a white tractor-trailer crossing the highway against a bright sky
FURTHER READING AND RESOURCES
HDR Libraries

- **OpenEXR**
  - Developed by Industrial Light & Magic
  - I/O for .exr format
- **PFSTools**
  - Command line tools and C/Matlab libraries
  - HDR I/O, tone mapping, viewing utilities
- **Matlab HDR Toolbox**
  - HDR functions for tone mapping, inverse tone mapping and utilities
- **Piccante**
  - Open source C++ library
  - HDR image processing
HDR Applications

- **Luminance HDR**
  - Based on PFS tools with QT interface
  - Cross-platform
  - HDR merging, tone mapping
HDR Applications

- **HDRshop**
  - Developed by Paul Debevec

- **Picturenaut**
  - Developed by Christian Bloch
  - Image based lighting, HDR Panorama creation

- **HDREfex Pro**
  - Plugin for Photoshop/Lightroom/Aperture
  - Creative photo effects

- **Photomatix**
  - HDR alignment and merging
HDR Content

- HDR Photographic Survey
  - Almost 100 HDR photographs
  - Radiometrically calibrated
  - Colorimetric measurements for several images
HDR Content

- **Stuttgart HDR Video Dataset**
  - Videos of challenging scenes with cinematic staging
  - Captured using mirror rig – 2 exposures
References

References – Perception

References – Capture

• HDR Capture / Sensors

• Ghost removal:
References – Tone mapping

- **Video Tonemapping**
References – Tone mapping

- **Inverse Tonemapping**
References – Color Management

• Color Spaces

• Color Appearance Modeling
References – Color Management

- **Gamut Management**
References – Video Compression

References – Display

- Kunkel, T., Daly, S., Miller, S., and Froehlich, J.,., 2016. Perceptual design for high dynamic range systems. High Dynamic Range Video: From Acquisition, to Display and Applications.
References – Quality of Experience


References – Quality of Experience


