3D image and video compression standards: past, present, future

Prof. Dr. Touradj Ebrahimi
EPFL
Introduction

- Evolution of cinema and television towards greater realism
  - “We went from standard definition to high definition, and [3D] is the natural next step.“ (Jim Spinella, HDlogix)

- Major factors for success
  - Affordable 3D technology for pros and consumers
  - Attractive content considering the new media
  - Improved quality of experience with respect to traditional media
Quality of Experience is the Key

- "As much as it pains me to say this - I love 3-D, I really do - these films are unpleasant to watch."

- "In any event, the 3D experience doesn’t at all feel natural, much less immersive."

- "I don't get motion sickness from a car or sea yet 3D movies tend to give me headache."

- "The 3D trend is annoying…what’s so bad about a really beautiful 2D composition? Even the best 3D still darkens the picture and muddies the color ever so slightly"

- "The 3D glasses offered to me ruined my movie-watching experience to say the least. They were uncomfortable to wear as they only had a one-size-fits-all."
3D perception cues

Accommodation

Binocular Disparity

Pictorial Cues

Motion Parallax

~10^{-1}m  ~10^1m  ~10^2m  ~10^3m  + inf
History of 3D

- **1840**: Invention of stereoscopy and stereoscope by C. Wheatstone
- **1890**: First patent for 3D motion pictures using stereoscope
- **1915**: First 3D footage in cinema using anaglyph glasses
- **1922**: Invention of „Teleview“ a shutter based technique
- **1936**: First demonstration of polarization based projection
- **1952**: Golden era of 3D movies due to invention of television
- **1961**: Single film solution „Space-Vision 3D“ using polarization
- **1980**: IMAX 70mm projectors for non-fiction short films
- **2003**: First full length 3D feature film for IMAX screens by J. Cameron
- **2004**: Animation „Polar Express“ makes 14 times more revenue in 3D than 2D
- **2009**: Movie „Avatar“ with a budget of 237 M$ becomes the biggest success of all times
3D Video coding approaches

- Simulcast
- Stereo (MPEG-4 AVC Stereo SEI)
- Mixed Resolution Stereo
- Video+Depth (MPEG-C Part 3)
- Multiview video (MPEG-4 MVC)
- Multiview video+depth (3DV)
Multiview video coding (MVC)

- Exploits redundancy between views using inter-camera prediction to reduce required bit-rate
Multiview + Depth coding

- Disconnecting the 3D video representation/coding from the captured video representation, and the displayed video representation
- Particularly interesting for auto-stereoscopic displays
- The current focus of MPEG 3D video coding standardization

**Multimedia Signal Processing Group**
**Swiss Federal Institute of Technology**
3DV test campaign

- **MPEG Call for Proposal (CfP) on 3D video coding (2011)**
  - “to define a data format and associated compression technology to enable the high-quality reconstruction of synthesized views for 3D displays”
  - Target: stereoscopic and auto-stereoscopic multi-view displays

- **Formal subjective quality evaluation test campaign (3DV tests)**
  - MPEG invited the European COST Action QUALINET
    - 12 evaluation laboratories *from around the world*
Test Scenarios and Requirement on Submissions

- **Test classes**
  - Class A: 1920x1088p@25fps
    Poznan Street, Poznan Hall2, Undo Dancer, GT Fly
  - Class C: 1024x768p@30fps
    Kendo, Balloons, Loverbird1, Newspaper

- **Test scenarios**
  - 2-view: refers to the 2-view input configuration
  - 3-view: refers to the 3-view input configuration

- **Test categories and anchors**
  - AVC-compatible: MVC encoder JMVC 8.3.1
  - HEVC-compatible: HEVC encoder HM 2.0

- **Coding conditions**
  - 4 different bit rates for each sequence, for each test scenario, for each test category
  - Pre-processing only for data format conversion (ex: down-sampling)
  - No optimization or processing steps using non-automatic means
Timeline and Responses to Call for Proposals

- **Timeline**
  - 2011/01/28: Draft CfP (public)
  - 2011/04/15: Final CfP
  - 2011/09/01: Submission of decoded test material
  - 2011/10/01: Subjective assessment starts
  - 2011/11/01: Submission of documents
  - 2011/11/26-12/02: Evaluation of proposals at 98th MPEG meeting

- **22 submissions**
  - Both AVC-compatible and HEVC-compatible

- **Test Model (TM) = best proposal + tools from other submissions**
  - AVC: Nokia (26% rate reduction for 2-view and 35% for 3-view)
  - HEVC: HHI (54% rate reduction for 2-view and 63% for 3-view)

- **Core Experiments (CE): investigate best proposed tools**
  - 10 CEs in AVC-compatible
  - 11 CEs in HEVC-compatible
12 evaluation laboratories
- Each laboratory was assigned a certain number of test sessions (stereoscopic, autostereoscopic, or both)
- 18 naïve viewers per test sequence

Dry runs to test common HW and SW set up
- Same monitors (native resolutions of 1920x1080 pixels)
  - 46” Hyundai S465D polarized stereo monitor
  - 52” Dimenco BDL5231V autostereo monitor
- Same evaluation methodology (screening, training, GUI, scoring sheets, etc…)
- Similar test room configuration
  - Max 3 (5) viewers seated in front of the stereoscopic (autostereoscopic) display, at 2.3m (3.5m) distance
  - Controlled light system (6500K color temperature)
  - Ambient luminance at 15% of max screen luminance
  - Each lab reported upon monitor calibration settings, gender and age statistics of the sample of viewers
Evaluation methodology

**Double Stimulus Impairment Scale (DSIS) evaluation**

“Rate the quality of each stimulus B, keeping in mind that of stimulus A”

**11-grade numerical categorical scale**
- 10: highest quality (i.e. test sequence indistinguishable from the reference)
- 0: lowest quality

**Basic test session:** 24 test pairs + 3 dummy pairs + 1 ref vs ref pair

- 16 sessions for Class A 2 views
- 16 sessions for Class B 2 views
- 16 sessions for Class A 3 views
- 32 sessions for Class B 3 views (include central view selection and random view selection)

16 sessions for Class B

Multimedia Signal Processing Group
Swiss Federal Institute of Technology
Overlapping data

- **4 laboratories:** NTNU, EPFL, Acreo, UBC
  - Class A 2-view stereo: 4 overlapping sessions (EPFL - UBC)
  - Class A autostereo: 4 overlapping sessions (EPFL - UBC)
  - Class B 2-view stereo: 8 overlapping sessions (NTNU - Acreo)
Scatter plots (EPFL – UBC)

**Autostereo data**

**Stereo data**

Pear: 0.98777  Spear: 0.98333

Pear: 0.95821  Spear: 0.96085
Scatter plots (NTNU – Acreo)

Stereo data
Core Experiments

AVC-compatible:
- View synthesis prediction for texture and depth
- Depth-based prediction
- Depth representation and coding
- Depth intra prediction without inter-component prediction
- Depth range compensation for inter/inter-view prediction
- In-loop depth resampling
- RD optimization through view synthesis distortion
- Global depth-and-view prediction
- Texture-based prediction for depth coding
- Depth In-loop Filtering

HEVC-compatible:
- View Synthesis Based Prediction for Texture
- View Synthesis Based Prediction for Depth
- Motion Parameter Prediction and Coding (independent of Depth)
- Transform Coding for Depth
- In-Loop Filtering for Depth
- Prediction Parameter Coding for Motion Data and Intra Prediction Modes
- Coding of Quantization Parameters
- Component Extraction
- Prediction Structures for Inter-view Prediction
- Modified Distortion Measure for Depth Coding
- View Synthesis
Standardization tracks considered

- MVC-compatible extension including depth
  - No block-level changes to AVC/MVC syntax and decoding processes
  - Add high-level syntax to enable efficient coding of depth data
  - Final Draft Amendment (FDAM): 2012/10

- AVC-compatible video-plus-depth extension
  - Change syntax and decoding process for non-base texture view and depth maps at block level
  - Expected coding efficiency improvements: 30-40% w.r.t. AVC/MVC
  - FDAM: 2013/07

- HEVC 3D extensions
  - Expected coding efficiency improvements: 40-60% w.r.t. HEVC (which itself is expected to achieve 50% rate reduction compared to AVC)
  - FDAM: 2014/01
Beyond standards

- **Stereo/Multiview image compression**
  - MPO, PNS, JPS
  - File formats based on JPEG or PNG compression

- **3D beyond binocular cues**
  - Holographic representation is still quite far-fetched
  - Holoscopic representation is more mature and probably more feasible