Coding of Stereoscopic and 3-Dimensional Video
Tutorial at IWSSIP 2012

Prof. Dr.-Ing. Jörn Ostermann
Leibniz Universität Hannover
Germany

Institut für Informationsverarbeitung
Acknowledgements

MPEG, Barry Haskell, Jens Ohm, Karsten Müller, Vittori Baroncini, Anthony Vetro, Richard Craig-McFeely, David A. Smith, Michael M. Bronstein, Yi Shi & Saul Rodriguez, David Wood, Sanggil Lee, F. Chen, H. Azari, Marc Schubin, Elliott Ellis, David Metge, Paul Verge, Romain Ziba, Marek Domanski, Masayuki Tanimoto, Dong Tian, Po-Lin Lai, Patrick Lopez, Cristina Gomila
If you experience fatigue or discomfort while viewing 3D images, cease use immediately. A ten-minute break is recommended about once every half hour. Switch to 2D immediately if 3D images still appear double after you have adjusted parallax. Individuals with a history of photosensitive epilepsy or heart disease or who are unwell or suffering from fatigue, insomnia, or the affects of alcohol should refrain from viewing 3D images. 2D display is also recommended for young children (up to the age of about six) whose visual system is still maturing and for individuals with notable differences in vision between their two eyes, who may find it difficult or impossible to observe the 3D effect. Viewing 3D images while in motion may cause fatigue or discomfort.

(example from Fujifilm 3D camera information)
Dimensions of Media

Radio

Black-and-white television

Color television

3D video
What Is 3D?

- **Motion**
  - what just happened?
- **Sound**
  - what’s she saying?
- **Color**
  - red dress? army green?
- **HD**
  - earring or blood?
- **3D**
  - depth information?
  - or just depth sensation?
Conventional 2D TV

You see exactly what the camera shot

Video: Bullettime
You see exactly what the **two** camera shot

Three-dimensional depth perception of the scene
Free viewpoint TV (FTV or 3DV)

Interactive selection of viewpoints

Video: Bullettime

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Augmented reality

Place 3D generated objects in the scene

Video: Japanese TV
Evolution

1851
Brewster streoscope sold in London

1891
Anaglyph is invented

1952
First color 3D movie premiere

Today
3D content available
User-grade 3D displays 3DTV broadcast

1922
First 3D movie premiere

Early 2000s
Commercial 3D monitors appear

1800 1820 1840 1860 1880 1900 1920 1940 1960 1980 2000

1838
Wheatson explains binocular vision

1839
Talbot invents photographic process

1928
First television broadcast in USA

1995
First IMAX 3D fiction movie

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
History of 3D Filming

- 1950s: 3D Cinema seen as saviour for movie theaters
  - Provided “exciting” alternative to new television services
- CES 2010: 3DTV seen as saviour for television set manufacturers
  - Provides “exciting” alternative to HDTV services

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
How likely are you to buy a 3DTV in the next 12 months?
(Base: Those interested in purchasing a new TV set in the next 12 months-n=142)

On a scale of 1 (not at all likely) to 5 (very likely)

Source: CTAM & The Nielsen Company
Predictions: 3D in the Home

• You will buy one
• You will not use it
  ○ “Showroom feature”
  ○ LASIK: $2k to take glasses off
  ○ 3D TV: $2k to put glasses back on
Challenges

- Representation (computer vision)
- Coding
- Rendering (computer vision/graphics)
Contents

• Camera
• Displays
• Human Visual System
• Coding Concepts of Multiview Video
• 3DV
• Future Standards
Depth Perception

- **Vergence:** Eyes directed towards object

- **Accommodation:** Eyes focus the lens according to distance of object
  - Limited depth of focus
  - Small sharp area
• Disparity: A point on an object appears displaced in the other image.
• Motion parallax: Images change depending on motion of object and viewer.
$\frac{x}{F} = \frac{X}{Z}, \quad \frac{y}{F} = \frac{Y}{Z},$

$x = F \frac{X}{Z}, \quad y = F \frac{Y}{Z}.$
Stereo Camera with Parallel Axis

\[
\frac{x_l}{F} = \frac{X + 0.5B}{Z} \quad \frac{x_r}{F} = \frac{X - 0.5B}{Z}
\]

Disparity: \[x_l - x_r = \frac{1}{Z} FB\]
Stereo Camera with Parallel Axis

Disparity: \[ x_l - x_r = \frac{1}{Z} FB \]

Red ball closest to camera
Stereo Camera with Converging Axis
• Keystone-Effect
  – Requires undistortion
• 3D is based on stereoscopic content
  – 2 views recorded from each scene:

  3D Stereoscopic Camera
  Content for the right eye
  Content for the left eye
  Two lenses/cameras fixed close together
  No crosstalk!
3D Parallel Filming

- One Camera per eye
  - ~63mm apart

- Minimal Distortions

- Fixed viewing focal point

- Easier to view

- Can Be Less Realistic
3D Converging Filming

- One Camera per eye
  - ~63mm apart
- Keystoning Distortions
- Variable Focus Point
- Can provide best picture
- But eye strain for near/far objects due to divergence
3D Convergence Filming

- Provides Potentially Best Picture Quality
- Keystoning distortion may need post production correction
- Focussing on Near and Far objects may cause eye strain
Free-View 3D Capture
- Capture the scene with multiple cameras closely located
- Transition between different views
- If the cameras are dense enough, the transition will be smooth
- Used in QuickTime VR
- Only existing viewpoints
Stereo cameras

IMAX 3D camera  NASA Mars rover  Multiview camera array (Carnegy Mellon)

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Stereo Camera
Stereo Camera
New 3D TV camera rig developed by KBS

Left eye camera

Right eye camera
• Left and right image appear shifted
• Stereo enables humans to perceive shape
• Camera
• **Displays**
• Human Visual System
• Coding Concepts of Multiview Video
• 3DV
• Future Standards
The Real World on a 2D Screen

Convergence

$Z_D$

$3D$ Objects

$3D$ Display

Accommodation

$Z_w$

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Eye-Glasses Stereoscopic

- Left: polarizing glasses
- Center: red-blue anaglyph
- Right: PC shutter-glasses

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Passive 3D TV: Anaglyphic Glasses

Right eye content tinted red
Left eye content tinted cyan

Each Frame contains spatially interleaved left and right content

• **Advantages:**
  – Wide Viewing Angle
  – Lightweight and cheap glasses
  – Can be used with any type of Display
  – New complex colour filters provide excellent quality

• **Disadvantages:**
  – Poor Colour Fidelity with Red/Cyan Filters
  – More accurate/expensive displays for new complex filters
Passive 3D TV Displays: Polarizing Glasses

- **Advantages:**
  - Wide Viewing Angle
  - Lightweight, cheap and stylish glasses
  - Can be used with projection systems

- **Disadvantages:**
  - Special polarizing displays needed
  - Reduced vertical resolution on current LCD/Plasma TVs (1920 x 1080i)

```
100/120Hz Frame Rate
```

Right Eye content
Clockwise polarized by TV

Left Eye content
Anticlockwise polarized by TV

```
100/120Hz Frame Rate
```

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Active 3D TV Displays: Shutter Glasses

- **TV Synchronizes shutter via RF or IR link**

100/120Hz Frame Rate

- Each frame contains only right eye or only left eye content

**Advantages:**
- Wide Viewing Angle
- High Picture Quality on Any Display

**Disadvantages:**
- LCD shutter response time can limit frame rates
- Expensive battery powered glasses

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Do we need glasses?
1950s: 3D Projection Systems

Autostereoscopic “Ivanov” with Raster Grating (Russia, 1940)

Anaglyphic Glasses

Polarizing Glasses
- Lenticular lens sending parts of the image to different eyes
- No glasses needed (autostereoscopic)
- Can be attached to a legacy monitor
- User-grade commercial products available from Sharp, Philips, etc.
- Viewing angle is a challenge
Lenticular
3DTV: Autostereoscopic Displays

- Array of lenses fixed to display
  - Large CRT, Plasma and LCD

- Advantages:
  - No glasses required
  - Multiple Views possible
  - Viewer Tracking possible

- Disadvantages:
  - Reduced Display Resolution
  - Viewing Angle very Critical
  - Cross talk

- October 2009: New 3D OLEDs from 3M
  - Wider viewing angle
  - Lightweight small displays e.g. Nintendo 3DS
Free-View Display

- Auto-stereoscopic (5 - 200 views)
  - 5-view display by ARCELIK
  - 45-view display by Holographica
3D TV: Head Mounted Displays

- **Advantages:**
  - No dependency on the viewing angle
  - No cross talk

- **Disadvantages:**
  - No Communal Experience
  - Expensive, heavy battery powered glasses
Ghosting (Crosstalk)

- Glasses
- Display
- Synchronization
- Viewing Angle
- Orientation
- Viewing Position
- Inability to Fuse

Shot through active-shutter glasses for 3D TV review
3D-Image Presentation Methods

- **Non-Stereoscopic Methods**
  - Holography
  - Volumetric (3D Pixels)
  - Laser-Beam Scanning
  - Psychological

- **Stereoscopic Methods**
  - Eye-Glasses Stereoscopic
    - Color filters
    - Polarizing filters
    - shutter glasses

- **Autostereoscopic**
  - Display-Type
    - Lenticular
    - Parallax Barrier
    - Integral Photography (IP)
    - Grating Array Plates

- **Projection-Type**
  - Fresnel Lenz
  - Holographic Screen
  - Spherical Mirror
Contents

- Camera
- Displays
- **Human Visual System**
- Coding Concepts of Multiview Video
- 3DV
- Future Standards
Shape from stereo

- Closer objects have larger parallax
- Depth recovered from parallax (disparity) between corresponding points
- Correspondence problem
- Can be generalized to multiple views

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Depth Perception

• **Vergence**: Eyes directed towards object

• **Accommodation**: Eyes focus the lens according to distance of object
  - Limited depth of focus
  - Small sharp area

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Vergence-Accommodation Conflict

from Prof. Martin Banks,
Visual Space Perception Lab
University of California – Berkeley

3.2 meters (10.5 feet) ok to $+\infty$

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Why We Can See the Stereo TV?

- The relation of the depth and the disparity length

![Graph showing the relation between apparent depth of stereoscopic image and disparity length on the screen.](image)
Countering V-A Conflict

In-Three “3D in the Home”
http://www.in-three.com/3DintheHomev2.html
Countering V-A Conflict

In-Three “3D in the Home”
http://www.in-three.com/3DintheHomev2.html
Which One Is Better, 2D or 3D?

- Comparison of psychological effects between 2-D and 3-D images

  - Sensation of Power
  - Total picture quality
• Objective evaluation of Psychological Effect
• The body sway of the viewer is tracked, measured and analyzed.
Miniaturization

Image position relatively easy to change; image content much harder to change.

*Elan Valley Miniature* (tilt-shift technique based on blur)
http://www.flickr.com/photos/frosted_peppercorn/481102393

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Which One Is Better, 2D or 3D?

• **Eye Fatigue**

  – Geometry, luminance and chrominance differences between the right and left images
  – Parallax that is very large or that rapidly changes spatially or temporally
  – Inconsistency between accommodation and convergence

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Chart 1. The average scores of subjective evaluation for the indoor stereoscopic scene
Chart 2. The average scores of our subjective evaluation for the outdoor HD stereoscopic scene
Subjective evaluation for 3DTV watching

* Surveyed by RAPA (Korea Radio Promotion Association) in 2010

**3D feeling**
- Very high: 10.2%
- High: 1.2%
- Average: 10.7%
- Low: 35.9%
- Very low: 42.0%

**Quality**
- Very high: 8.8%
- High: 0.9%
- Average: 31.3%
- Low: 46.8%

**Eye fatigue**
- Very high: 16.7%
- High: 2.7%
- Average: 5.2%
- Low: 32.3%
- Very low: 43.1%

**Dizziness**
- Very high: 20.4%
- High: 2.4%
- Average: 2.6%
- Low: 27.1%
- Very low: 47.6%

* Surveyed by RAPA (Korea Radio Promotion Association) in 2010

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de

64
Eyes versus Stereo Camera

- 2 eyes
- Accommodation and vergence coupled
- Disparity
- Motion parallax

- 2 cameras
- No coupling
  - motors, interactivity, gaze tracking
- Disparity
- No
  - Motors, interactivity, gaze tracking
ITU-Generations (Profiles) of 3D-TV

• **First Generation 3D-TV.** Record two samples of the Object Wave (Amplitude, Frequency via primaries, No phase). Characteristic: depth, but nothing changes as it should when you move your head. Focusing doesn’t work.

• **Second Generation 3D-TV.** Record a large number of pairs of signals (multi-view). A ‘quantized’ horizontal Object Wave. Move your head, you get different views. Display resolution limitations today.

• **Third Generation.** Horizontal and vertical quantized Object wave.

• **Fourth Generation 3D-TV.** Record the entire Object Wave. Modulate the Object wave to make it ‘recordable’. Hologram does this. Massive bandwidth needed. Just like natural vision. No eyestrain.

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
• Camera
• Displays
• Human Visual System
• **Coding Concepts of Multiview Video**
• 3DV
• Future Standards
• **Transmission ready** (DVB, Blu-Ray, …)
• **Coding**
  – Compatibility, market introduction, …
  – Many existing standards
  – Rate proportional to no. of views?
Concepts applicable to:
MPEG-1, **MPEG-2**, MPEG-4 Part 2, **AVC|H.264**, HEVC

- Simulcast
- Side-by-Side
- MVC
- Video plus Depth
Simulcast

INPUT VIDEO 1 ➔ H.264/AVC encoder ➔ Bit stream ➔ Channel ➔ H.264/AVC encoder ➔ OUTPUT VIDEO 1

INPUT VIDEO 2 ➔ H.264/AVC encoder ➔ Bit stream ➔ Channel ➔ H.264/AVC encoder ➔ OUTPUT VIDEO 2
Simulcast
Frame Packing: Spatially Interleaved Views

- **Side-by-Side 960 x 1080i**
- **Over/Under 1920 x 540i**

- **Reduces resolution by up to 50%**
  - Side by Side can give better resolution with careful camera positioning

- **(Can Be) Fully Compatible with existing Encoder/Decoders**
  - 3D TV Set generates two separate frames from combined frame

- **Side-by-Side 960 x 1080i Adopted by BSkyB**
  - Services started April 2010 using existing set top box

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
### Side-by-Side: HDMI 1.4/1.3 Stereo-TV Formats

<table>
<thead>
<tr>
<th>Frame Packing</th>
<th>Side-by-Side (Half)</th>
<th>Top-and-Bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920</td>
<td>960</td>
<td>1920</td>
</tr>
<tr>
<td>L</td>
<td>R</td>
<td>L</td>
</tr>
<tr>
<td>1080</td>
<td>960</td>
<td>1080</td>
</tr>
</tbody>
</table>

- **1080p, 23.98 / 24 Hz**
- **720p, 59.94 / 60 Hz**
- **720p, 50 Hz**

<table>
<thead>
<tr>
<th>2 x HD</th>
<th>1 x HD</th>
<th>1 x HD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1080p</td>
<td>1080i</td>
<td>1080p</td>
</tr>
<tr>
<td>720p</td>
<td>1080i</td>
<td>720p</td>
</tr>
<tr>
<td>50 Hz</td>
<td>50 Hz</td>
<td>50 Hz</td>
</tr>
</tbody>
</table>
MVC: AVC Multi View Coding

- Defined as part of AVC/MPEG4/H.264
  - ISO/IEC 14496-10:2008 Amendment 1
• Inter-frame and Inter-view (disparity) differences
• Based on standard AVC/H.264 compression techniques
Comparison AVC - MVC

PSNR Y vs. bitrate, exit sequence, all views

Average PSNR Y, dB

Average bitrate, kbps

MVC

AVC

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Efficiency of Frame Packing using AVC

- Source: “H.264/AVC Stereo Video Compression Benchmarking”
- Subarna Tripathi, Emiliano Mario Piccinell, Davide Aliprandi

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Comparison AVC - MVC

Gain depends on number of views
- Up to 20%

50% rate increase compared to mono
Key Technologies for an Advanced 3D-TV System
C. Fehn, K. Hopf, B. Quante (Fraunhofer Inst. HHI)
2D Video plus Depth (V+D)

Image

Depth

3D Warp

Virtual view
– 1 or more videos plus depth map
2D Video+Depth

Conventional video (2D)  Depth (Z)

- 2D video and corresponding depth map
- Depth-component increases bandwidth only by 5-20%
- Used by Philips (WOW vx)
- User can adjust perceived depth for comfort
- No pickup for broadcast
Depth Map sent as auxiliary video stream

- MPEG-C part 3 (ISO23002-3)
- Depth Map computation difficult for natural scenes
Contents

- Camera
- Displays
- Human Visual System
- Coding Concepts of Multiview Video
- 3DV
- Future Standards
Stereoscopic displays
- Variable stereo baseline
- Adjust depth perception

Auto-stereoscopic N-view displays
- Wide viewing angle
- Large number of output views

Constrained Rate (based on distribution)

Limited Camera Inputs

Data Format

Left

Right

Data Format

3DV

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
3DV Evaluation

Data Format

Identical Data Format

Supl. Data

Codec

Supl. Data′

Synth.

V1  V2  V3

S1  S2  S3  SN

Input View Quality
PSNR of Video component
Bit Rate of entire data format

Synthesis Quality
Subjective Evaluation,
MOS of synthesized views

Evaluation

• One submission scenario, covering different data formats, compression, view synthesis
• Subjective 3D Evaluation of synthesized (random) stereo pairs / multi view
• Additionally, compression efficiency tested via automatic PSNR of video data and total overall bit rate (including all data format components)
Information for Proponents

• **Compatibility**
  – AVC-compatibility: partial fulfillment of requirements, if base view is AVC-coded, second (and third) view coded with some new approach
  – HEVC-compatibility: Indicate HM version number

• **Data rate**
  – All data and parameters, required by a proponent’s synthesis method to synthesized all views (according to test conditions) must be part of the bit stream
  – Compression of camera parameters

• **Related Software**
  – Depth estimation (DERS) and synthesis (VSRS) software ftp://ftp.merl.com/pub/avetro/3dv-cfp/
  – Stereo2Avi
  – Provision under GPL
• 3x3 block matching, depth recalculation using image segmentation and variable smoothing coefficient for graph cuts
\[
Z = \frac{1}{\frac{Y}{255} \left( \frac{1}{Z_{\text{near}}} - \frac{1}{Z_{\text{far}}} \right) + \frac{1}{Z_{\text{far}}}}
\]
View Synthesis Reference Software (VSRS)

- Boundary aware hole filling

- Synthesis from left and right
View Synthesis Reference Software (VSRS)

- Boundary noise removal
• **EE1: Depth estimation**  
  – Depth map refinement for one sequence required

• **EE2: View synthesis (extrapolation)**  
  – Several techniques and configurations have been tested  
  – Quality of extrapolated views not satisfactory in general
3DV Evaluation

- One submission scenario, covering different data formats, compression, view synthesis
- Subjective 3D Evaluation of synthesized (random) stereo pairs / multi view
- Additionally, compression efficiency tested via automatic PSNR of video data and total overall bit rate (including all data format components)
3DV CfP Timeline

- CfP at 95/96th meeting
- Evaluation prior to 98th meeting (11/2011)
- Call open to all technologies
  - AVC
  - MVC
  - HEVC
  - Mesh
  - Hybrid
Test Material

- **Test Classes**
  - Class C: 1024x768p 30fps: "Kendo", "Balloons", "Lovebird1", "Newspaper"

- **Test Scenarios**
  - 2-view: refers to the 2-view input configuration, one view coded, one view synthesized
  - 3-view: refers to the 3-view input configuration (one or two views synthesized; 28 views)
Test Categories

- **AVC-Compatible**: refers to submissions in which the compressed data format satisfy the requirement on forward compatibility with AVC [3]
- **HEVC-Compatible & Unconstrained**: refers to submissions in which the compressed data formats satisfy the requirement on forward compatibility with HEVC, or submissions without any compatibility constraints [3]
### Test Conditions

<table>
<thead>
<tr>
<th>Seq. ID</th>
<th>Test Sequence</th>
<th>2-view input</th>
<th>3-view input</th>
</tr>
</thead>
<tbody>
<tr>
<td>S01</td>
<td>Poznan_Hall2</td>
<td>7-6</td>
<td>7-6-5</td>
</tr>
<tr>
<td>S02</td>
<td>Poznan_Street</td>
<td>4-3</td>
<td>5-4-3</td>
</tr>
<tr>
<td>S03</td>
<td>Undo_Dancer</td>
<td>2-5</td>
<td>1-5-9</td>
</tr>
<tr>
<td>S04</td>
<td>GT_Fly</td>
<td>5-2</td>
<td>9-5-1</td>
</tr>
<tr>
<td>S05</td>
<td>Kendo</td>
<td>3-5</td>
<td>1-3-5</td>
</tr>
<tr>
<td>S06</td>
<td>Balloons</td>
<td>3-5</td>
<td>1-3-5</td>
</tr>
<tr>
<td>S07</td>
<td>Lovebird1</td>
<td>6-8</td>
<td>4-6-8</td>
</tr>
<tr>
<td>S08</td>
<td>Newspaper</td>
<td>4-6</td>
<td>2-4-6</td>
</tr>
</tbody>
</table>
## Rates

<table>
<thead>
<tr>
<th>Seq. ID</th>
<th>Test Sequence</th>
<th>2-view test scenario Bit rates (kbps)</th>
<th>3-view test scenario Bit rates (kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>R1</td>
<td>R2</td>
</tr>
<tr>
<td>S01</td>
<td>Poznan_Hall2</td>
<td>500</td>
<td>700</td>
</tr>
<tr>
<td>S02</td>
<td>Poznan_Street</td>
<td>500</td>
<td>700</td>
</tr>
<tr>
<td>S03</td>
<td>Undo_Dancer</td>
<td>1000</td>
<td>1300</td>
</tr>
<tr>
<td>S04</td>
<td>GT_Fly</td>
<td>1200</td>
<td>1700</td>
</tr>
<tr>
<td>S05</td>
<td>Kendo</td>
<td>400</td>
<td>500</td>
</tr>
<tr>
<td>S06</td>
<td>Balloons</td>
<td>320</td>
<td>430</td>
</tr>
<tr>
<td>S07</td>
<td>Lovebird1</td>
<td>375</td>
<td>500</td>
</tr>
<tr>
<td>S08</td>
<td>Newspaper</td>
<td>400</td>
<td>525</td>
</tr>
</tbody>
</table>
## View Synthesis

<table>
<thead>
<tr>
<th>Seq. ID</th>
<th>Test Sequence</th>
<th>View to Synthesize from 2-view test scenario (and stereo pair)</th>
<th>Views to Synthesize from 3-view test scenario (and stereo pair)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S01</td>
<td>Poznan_Hall2</td>
<td>6.5 (6.5-6)</td>
<td>All 1/16 positions between views 7 and 5 (6.125-5.875)</td>
</tr>
<tr>
<td>S02</td>
<td>Poznan_Street</td>
<td>3.5 (3.5-3)</td>
<td>All 1/16 positions between views 5 and 3 (4.125-3.875)</td>
</tr>
<tr>
<td>S03</td>
<td>Undo_Dancer</td>
<td>3 (3-5)</td>
<td>All 1/4 positions between views 1 and 9 (4.5-5.5)</td>
</tr>
<tr>
<td>S04</td>
<td>GT_Fly</td>
<td>4 (4-2)</td>
<td>All 1/4 positions between views 9 and 1 (5.5-4.5)</td>
</tr>
<tr>
<td>S05</td>
<td>Kendo</td>
<td>4 (4-5)</td>
<td>All 1/8 positions between views 1 and 5 (2.75-3.25)</td>
</tr>
<tr>
<td>S06</td>
<td>Balloons</td>
<td>4 (4-5)</td>
<td>All 1/8 positions between views 1 and 5 (2.75-3.25)</td>
</tr>
<tr>
<td>S07</td>
<td>Lovebird1</td>
<td>7 (7-8)</td>
<td>All 1/12 positions between views 4 and 8 (5.75-6.25)</td>
</tr>
<tr>
<td>S08</td>
<td>Newspaper</td>
<td>5 (5-6)</td>
<td>All 1/12 positions between views 2 and 6 (3.75-4.25)</td>
</tr>
</tbody>
</table>
Tool Categories

- **Texture coding**
  - Independent of depth
    - E.g., inter-view prediction of color view, inter-view prediction of motion parameters and residual data
  - Using depth data
    - E.g., view synthesis prediction, motion prediction, motion-compensated prediction, QP selection, inloop filter

- **Depth coding**
  - Independent of texture
    - E.g., depth modeling modes, weighted prediction
  - Using texture data
    - E.g., motion parameter inheritance, intra prediction
• **Huge effort**
  – 23 (12 + 11) submissions
  – 2-view and 3-view encodings
  – Evaluation on stereo and auto-stereoscopic displays
  – Major support from Qualinet and test labs from Europe, Canada, and Asia
  – 400 viewing subjects
Sample Subjective Test Result

Class A Sequence, 3-view test case
AVC-Compatible

Anchor at R3
Best Performing Proposal at R1
Class A Sequence, 2-view test case
HEVC-Compatible

Anchor at R3
Best Performing Proposal at R1
2 View Scenario

**2View S01**

- **MOS** vs **Bit Rate [kbit/s]**

- **AVC Anchor**
- **HEVC Anchor**
- **P18 (HEVC)**
- **P28 (AVC)**

- **58% (300kbps)**
- **55% (175kbps)**
- **25% (175kbps)**
- **27.5% (275kbps)**
- **40% (600kbps)**

---

Prof. Dr.-Ing. Jörn Ostermann  
ostermann@tnt.uni-hannover.de
2 View Scenario

![Graph showing MOS vs Bit Rate for different scenarios and bitrates.]

- 51% (350kbps)
- 44% (570kbps)
- 40% (320kbps)

Legend:
- AVC Anchor
- HEVC Anchor
- P18 (HEVC)
- P28 (AVC)
2 View Scenario

![Graph showing MOS scores against bit rate for different video encodings: AVC Anchor, HEVC Anchor, P18 (HEVC), P28 (AVC). The graph highlights MOS scores of 59% for 650 kbps and 52% for 380 kbps.](image)

2View S04

- AVC Anchor
- HEVC Anchor
- P18 (HEVC)
- P28 (AVC)

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
3 View Scenario

![Graph showing MOS vs Bit Rate for different scenarios.](image)

- **AVC Anchor**
- **HEVC Anchor**
- **P18 (HEVC)**
- **P28 (AVC)**

**3View S04**

- 28% (1100kbps)
- 54% (860kbps)

---

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
3D Video

• First assessment shows that inclusion of depth map coding could save 40+% compared to MVC, 60+% compared to HEVC simulcast (However, HEVC is already better by itself)
• HEVC Stereo could be more efficient than AVC Mono
• Rate grows with number of coded views

• One approach based on mesh-based decoder-side view synthesis
• **At least 3 different possibilities of standardization:**
  - MVC compatible (extending to carriage of depth maps, only extending some high-level syntax)
  - AVC base-view compatible (using specific tools for depth maps, and also utilizing them in coding other views from base view by specific tools)
  - HEVC compatible (extending HEVC to multi-view case and depth map coding)

• **Two more possibilities:**
  - Hybrid AVC/MVC-base + HEVC
  - Without depth map (generated at decoder side)
Contents

- Camera
- Displays
- Human Visual System
- Coding Concepts of Multiview Video
- 3DV
- Future Standards
MPEG-H: Standards for SmartMedia Technologies

- Richmedia
- Metadata
- Embedded CAS/DRM
- Embedded UI

Self-contained packaging

- User context
- Device capability
- Network QoS (Streaming/Download)

Context Awareness

- Advanced (Intuitive) User Interaction
- Controlling Environment
- Driving actuators

Active (Immersive) Playback

- Live update
- Mobility
- Social Networking
- Information-based hyperlinking

Connectivity Awareness

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
MPEG-H: Standards for SmartMedia Technologies

Smart Media

- Descriptive Composition Information
- Media Assets (Resources)
- Metadata
- Scripts or Executable

Interface for advanced creation & management

Interface for advanced delivery

Interface for advanced experience & interaction

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
MPEG-H: Initial Set of Standards

• **High Efficiency Video coding (HEVC)**
  – 2D Video
  – 3D Video (pure HEVC)
  – Scalability

• **Advanced audio coding**

• **MPEG Media Transport (MMT)**

• Graphics
• User Interfaces
HEVC: Assessment of Achievements

• Compression improvements compared AVC anchors in terms of gross “BD rate” based on PSNR are typically in the range of 40-45%.
• Gain is typically greater for high resolution and for low delay usage scenarios.
• Visual improvement seems to exceed PSNR-measured effects (perhaps by about 10%).
• Visual quality comparison is planned before the next meeting.
• Preliminary real-time implementations of software HEVC decoders were demonstrated.
### Random Access vs. Low Delay vs. All Intra

<table>
<thead>
<tr>
<th>Class</th>
<th>Random Access</th>
<th>Low Delay</th>
<th>All Intra</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A</td>
<td>43%</td>
<td></td>
<td>29%</td>
</tr>
<tr>
<td>Class B</td>
<td>44%</td>
<td>48%</td>
<td>26%</td>
</tr>
<tr>
<td>Class C</td>
<td>34%</td>
<td>41%</td>
<td>23%</td>
</tr>
<tr>
<td>Class D</td>
<td>32%</td>
<td>38%</td>
<td>18%</td>
</tr>
<tr>
<td>Class E</td>
<td></td>
<td>51%</td>
<td>29%</td>
</tr>
<tr>
<td>Average</td>
<td><strong>39%</strong></td>
<td><strong>44%</strong></td>
<td><strong>25%</strong></td>
</tr>
</tbody>
</table>
Potential other Standards for 3DV

- MPEG-2 plus HEVC
- AVC plus HEVC
- 3DV based on HEVC
- 3DV based on AVC
Future Challenges

• **Content**
  - Computer vision for depth
    - Transparent objects
    - Small objects
    - Non-rigid objects

• **Distribution**
  - Blu-Ray defined
  - Cinema: Stereo, no bandwidth concern

• **Early adopters**
  - Sports
  - Pay-TV
Final Remarks

• ...
• Opaque objects / transparent objects?
• What is the killer application for 3DV?
• How will the recording equipment become application friendly?
• ...

Prof. Dr.-Ing. Jörn Ostermann
ostermann@tnt.uni-hannover.de
Conclusions

• **Stereo TV arrived**
  – Legacy infrastructure
    ➢ MPEG-2
    ➢ AVC
  – BluRay
    ➢ MVC

• **3DV**
  – Representation
  – Computer Vision
  – Rendering

• **3D Geometry**
  – Synthetic Content
  – Games

• **Content and displays drive technology**
  – No holography