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

Caution

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

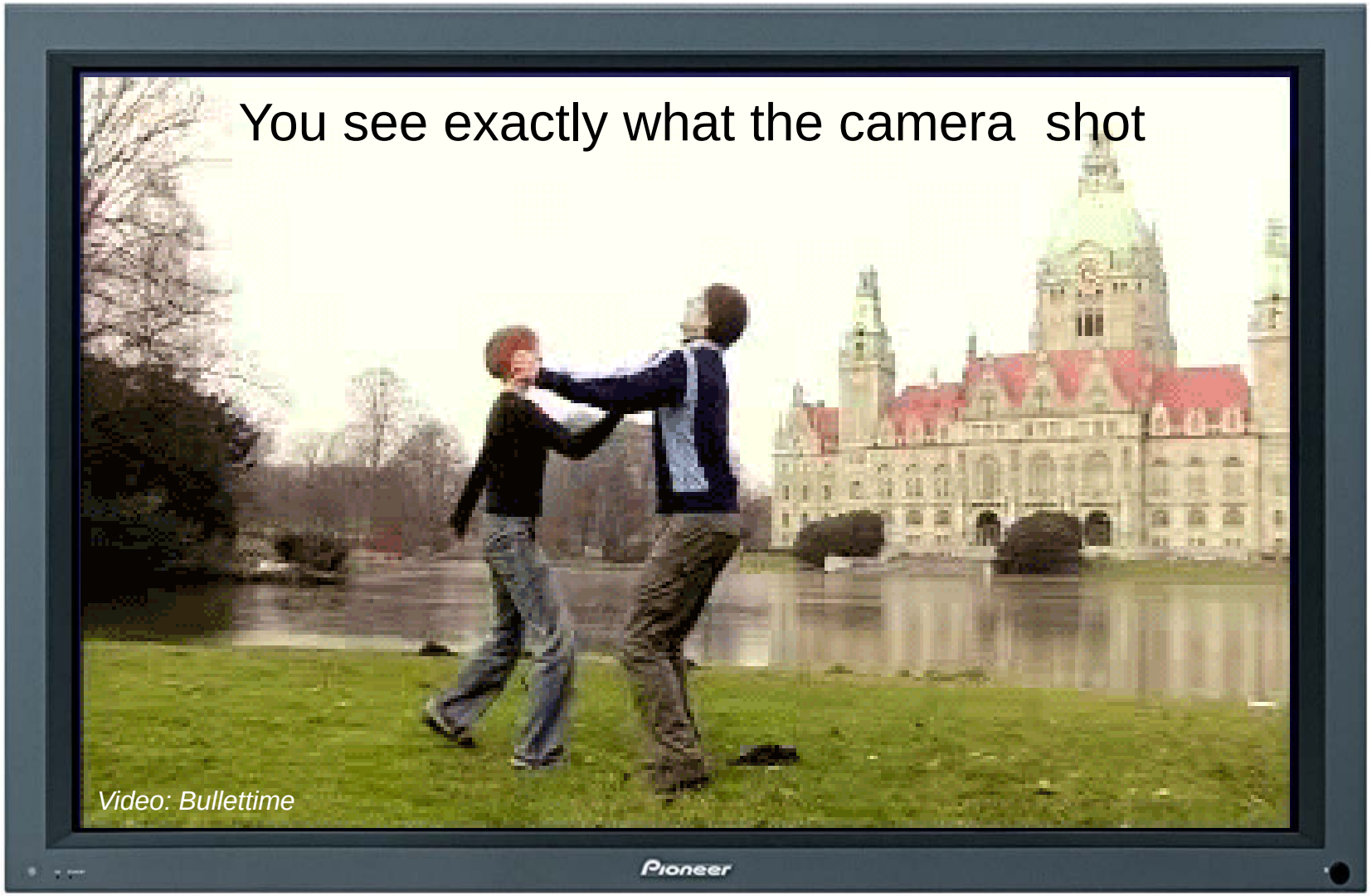
Color television

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



Stereoscopic TV

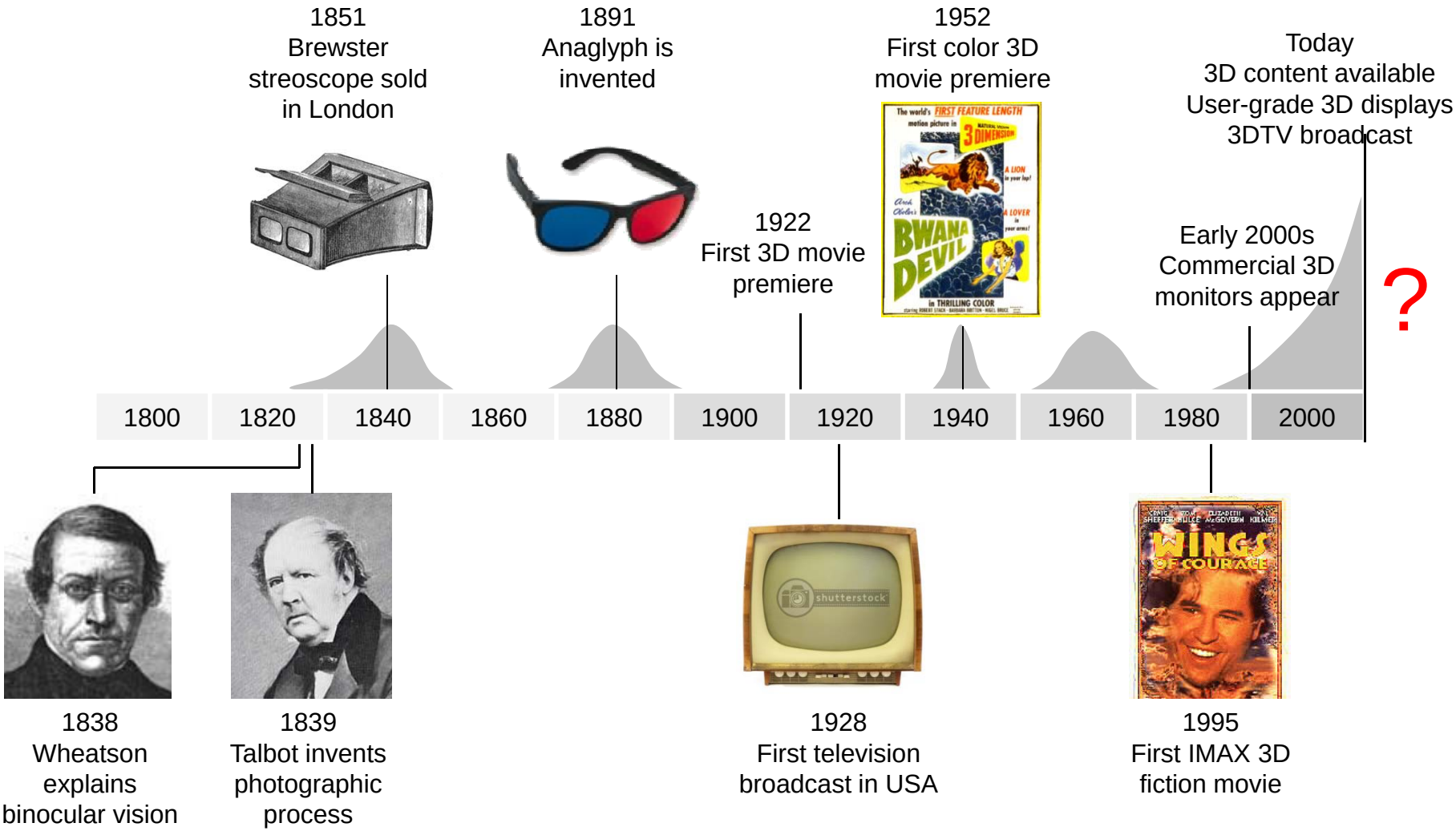
You see exactly what the two camera shot

Three-dimensional depth perception of the scene

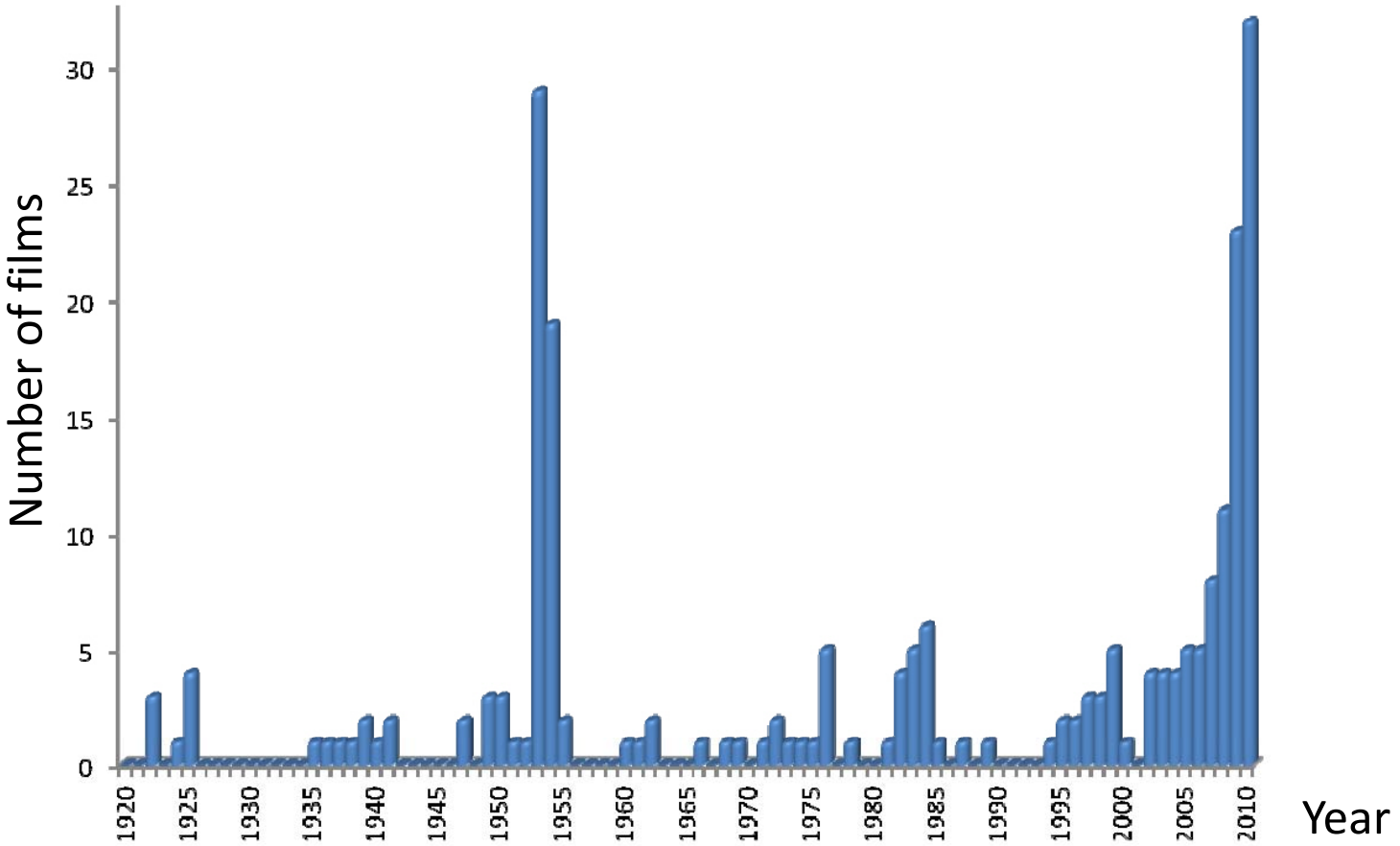
Free viewpoint TV (FTV or 3DV)

Augmented reality

Evolution

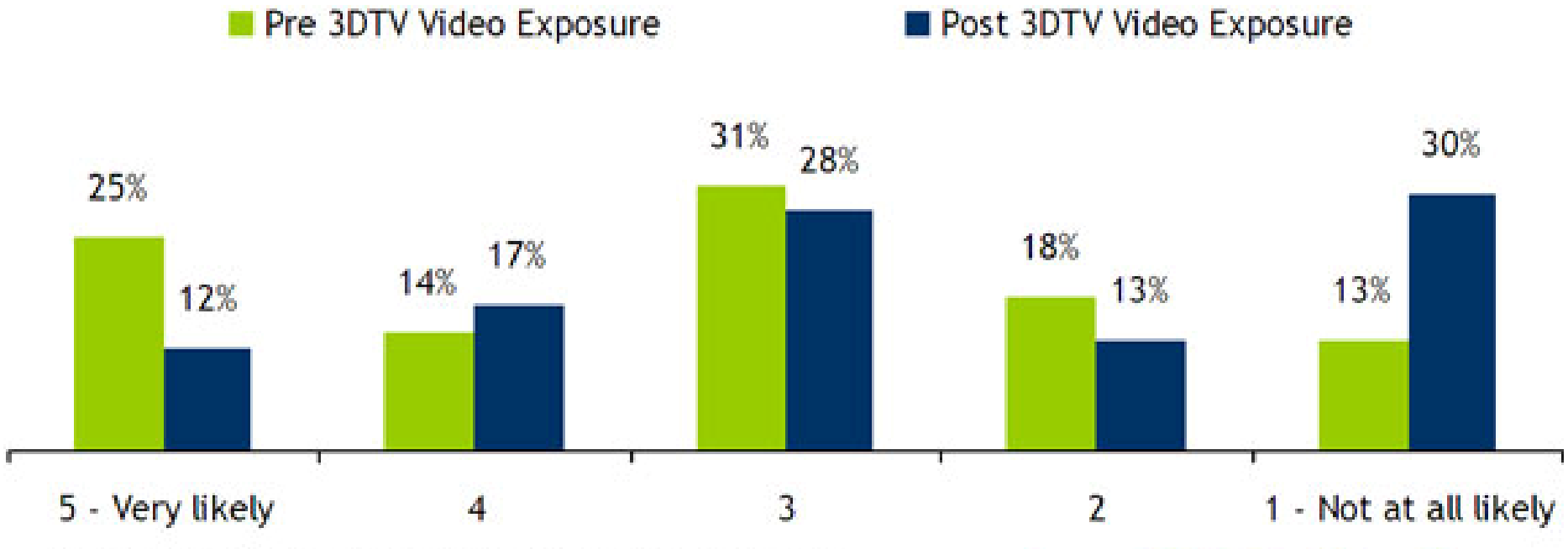

1851
Brewster
streoscope sold
in London
1891
Anaglyph is
invented
1922
First 3D movie
premiere
1952
First color 3D
movie premiere
Today
3D content available
User-grade 3D displays
3DTV broadcast
Early 2000s
Commercial 3D
monitors appear
?
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000
1838
Wheatson
explains
binocular vision
1839
Talbot invents
photographic
process
1928
First television
broadcast in USA
1995
First IMAX 3D
fiction movie

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

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)



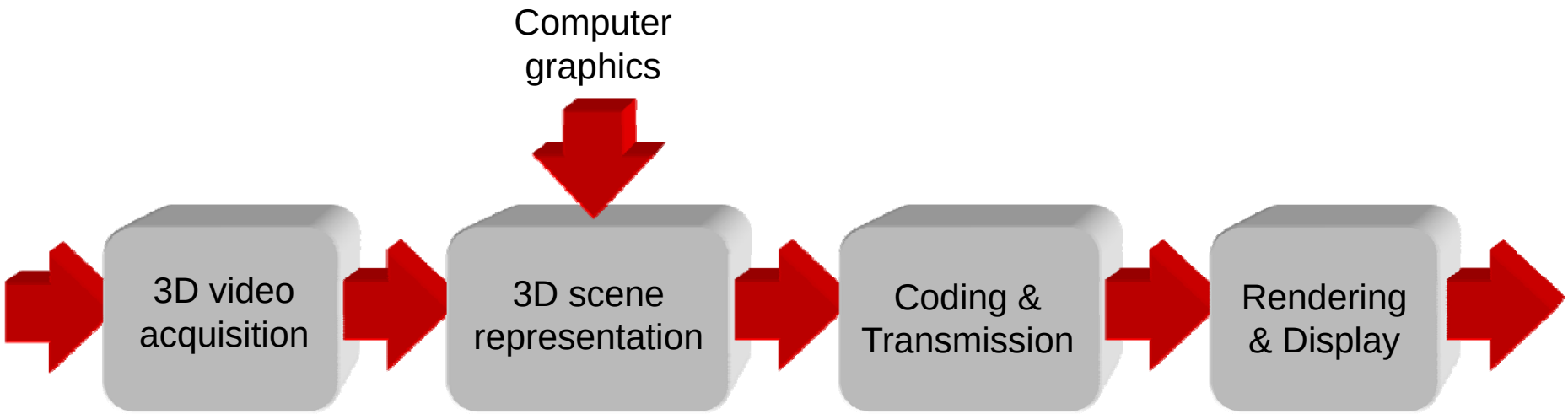
On a scale of 1 (not at all likely) to 5 (very likely)

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
  - 3DTV: $2k to put glasses back on

Challenges



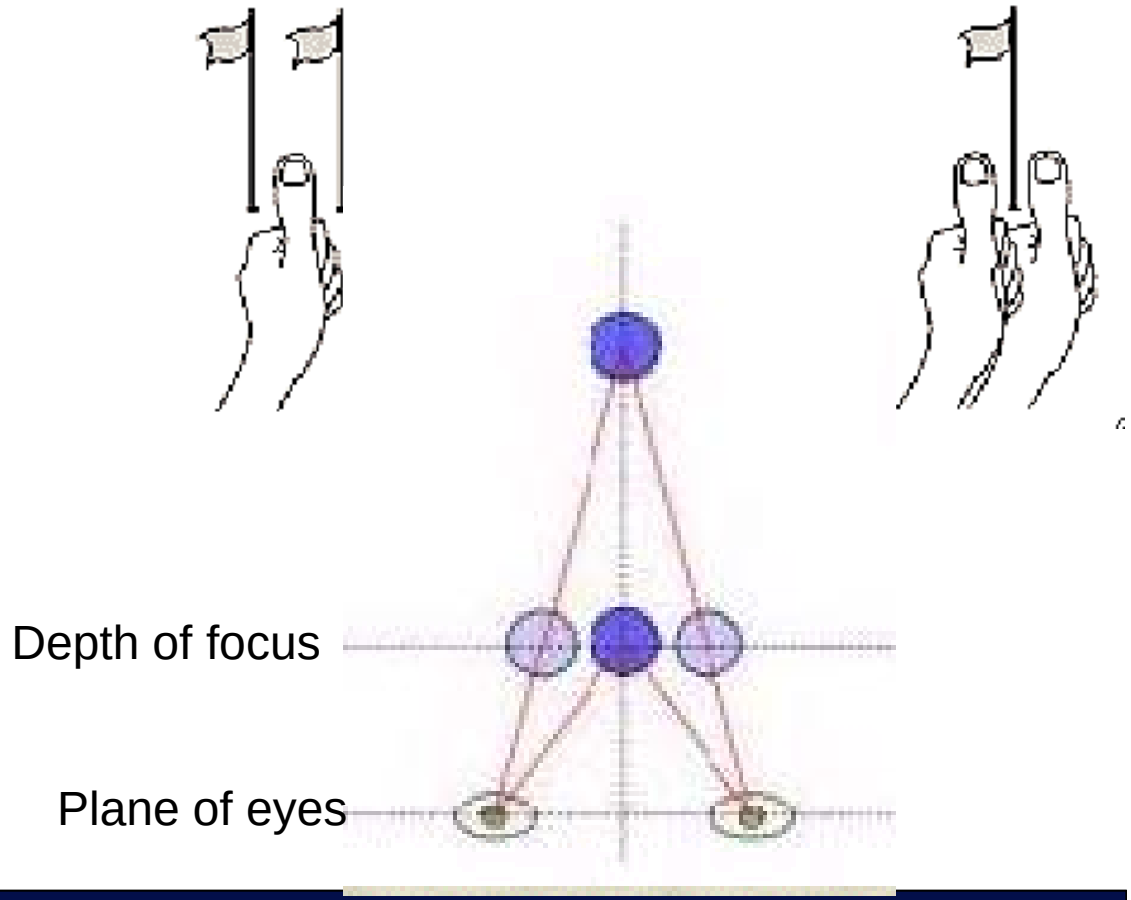
- Representation (computer vision)
- Coding
- Rendering (computer vision/graphics)

Contents

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

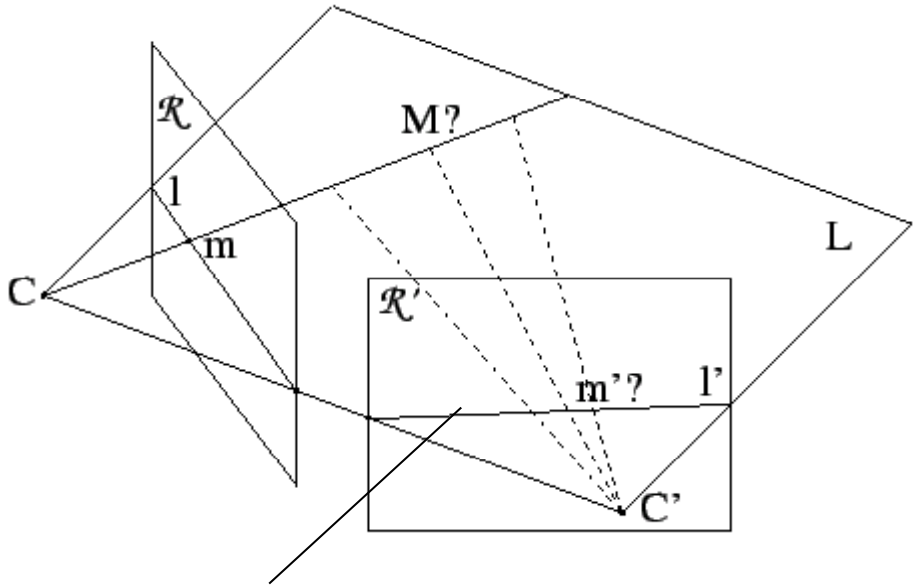
Depth Perception

- **Vergence: Eyes directed towards object**
- **Accomodation: Eyes focus the lens according to distance of object**
  - Limited depth of focus
  - Small sharp area



Depth Perception

- Disparity: A point on an object appears displaced in the other image.
- Motion parallax: Images change depending on motion of object and viewer.

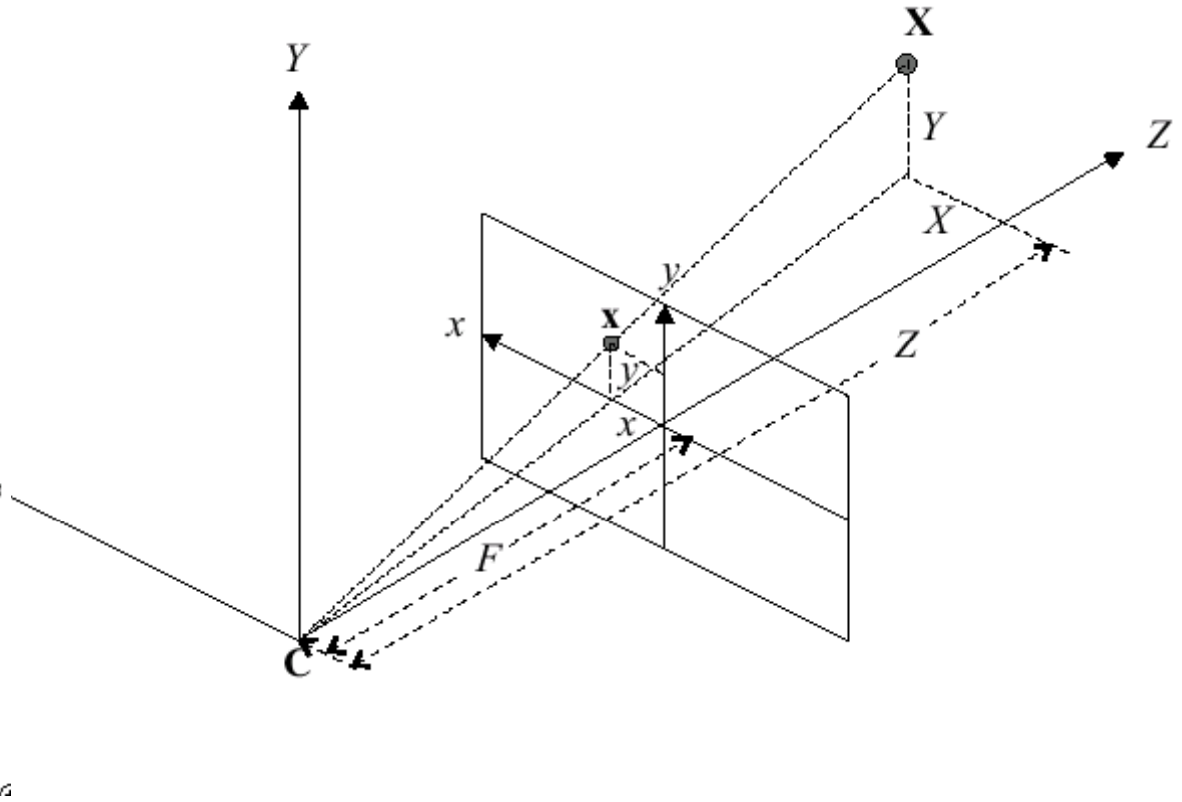


Epipolar-Line

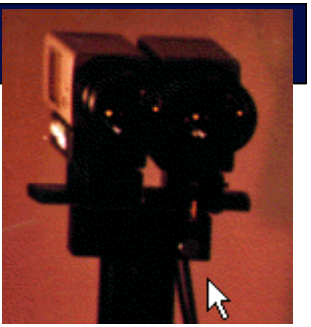
Pinhole Camera

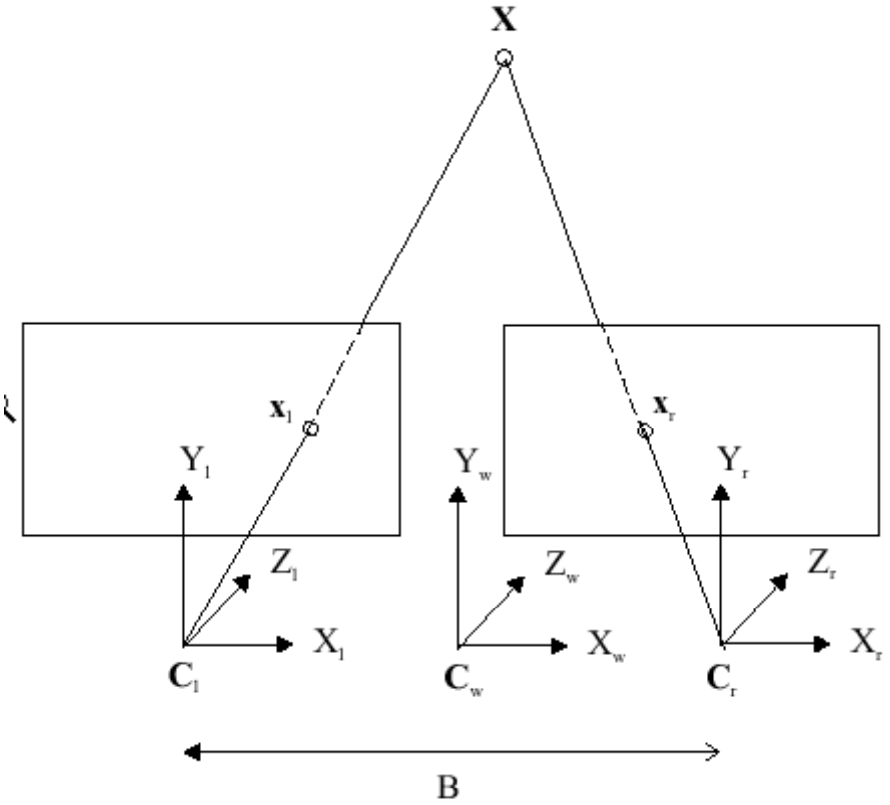
$$\frac{x}{F} = \frac{X}{Z}, \quad \frac{y}{F} = \frac{Y}{Z},$$

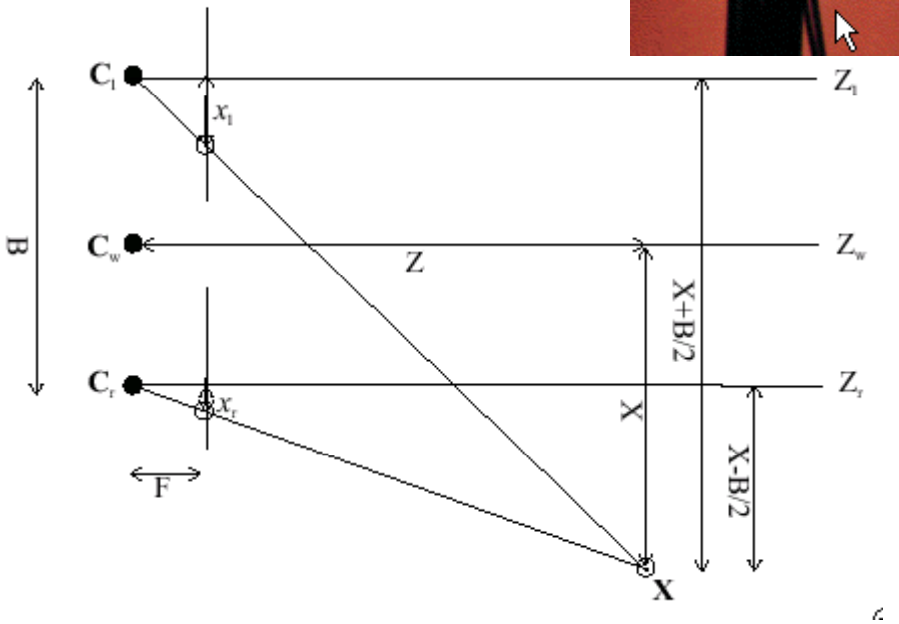
$$x = F\frac{X}{Z}, \quad y = F\frac{Y}{Z}$$



Stereo Camera with Parallel Axis



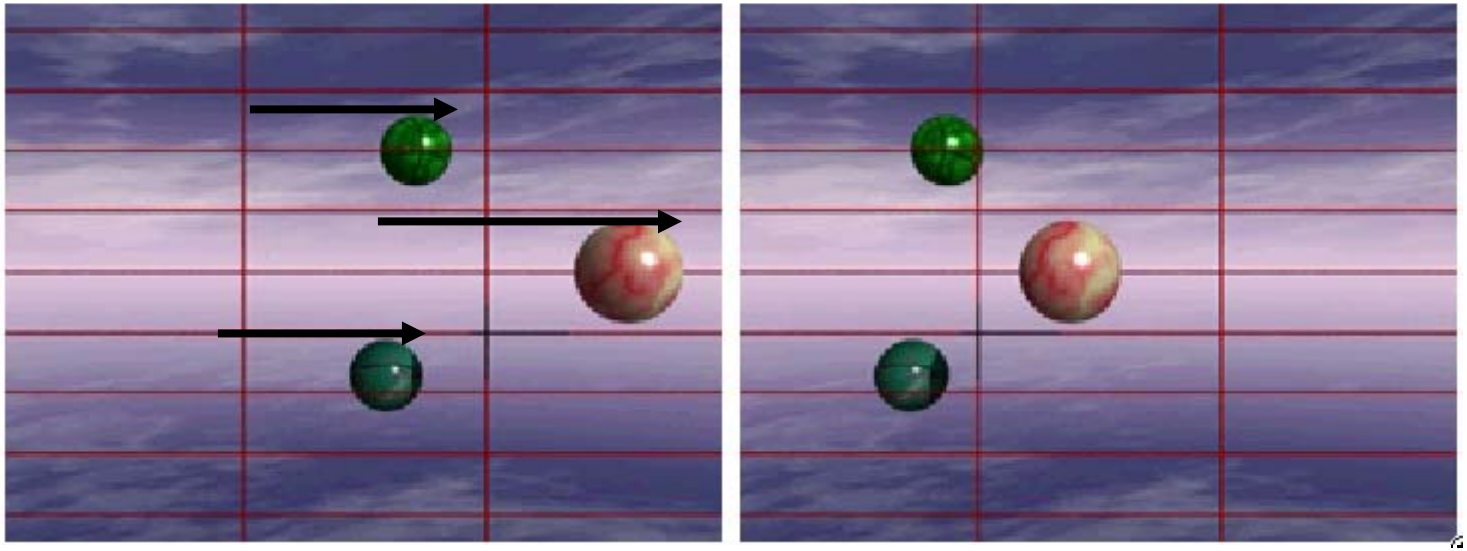




$$\frac{x_l}{F} = \frac{X + 0.5B}{Z} \qquad \frac{x_r}{F} = \frac{X - 0.5B}{Z}$$

$$\text{Disparity:} \qquad x_l - x_r = \frac{1}{Z} FB$$

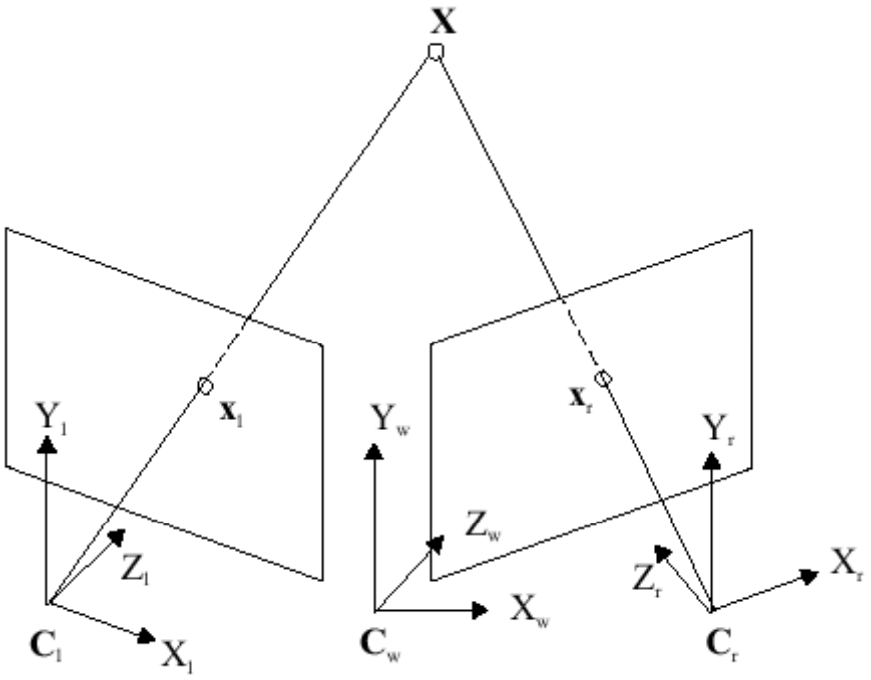
Stereo Camera with Parallel Axis



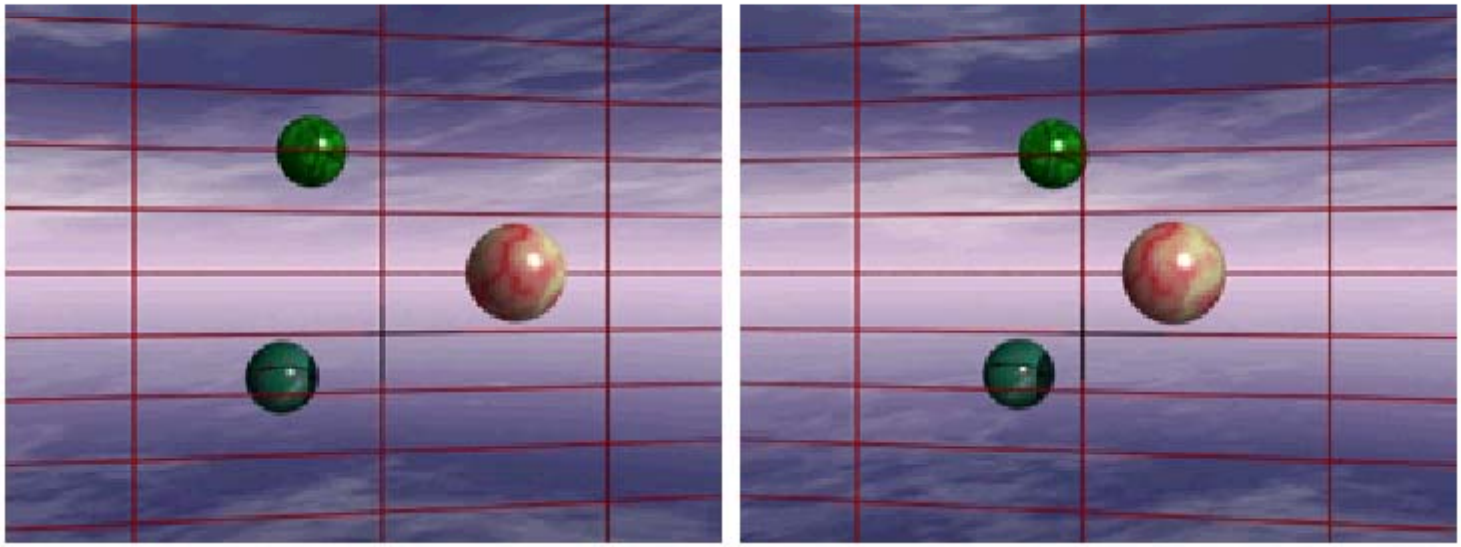
$$\text{Disparity}: \qquad x_l - x_r = \frac{1}{Z} FB$$

Red ball closest to camera

Stereo Camera with Converging Axis



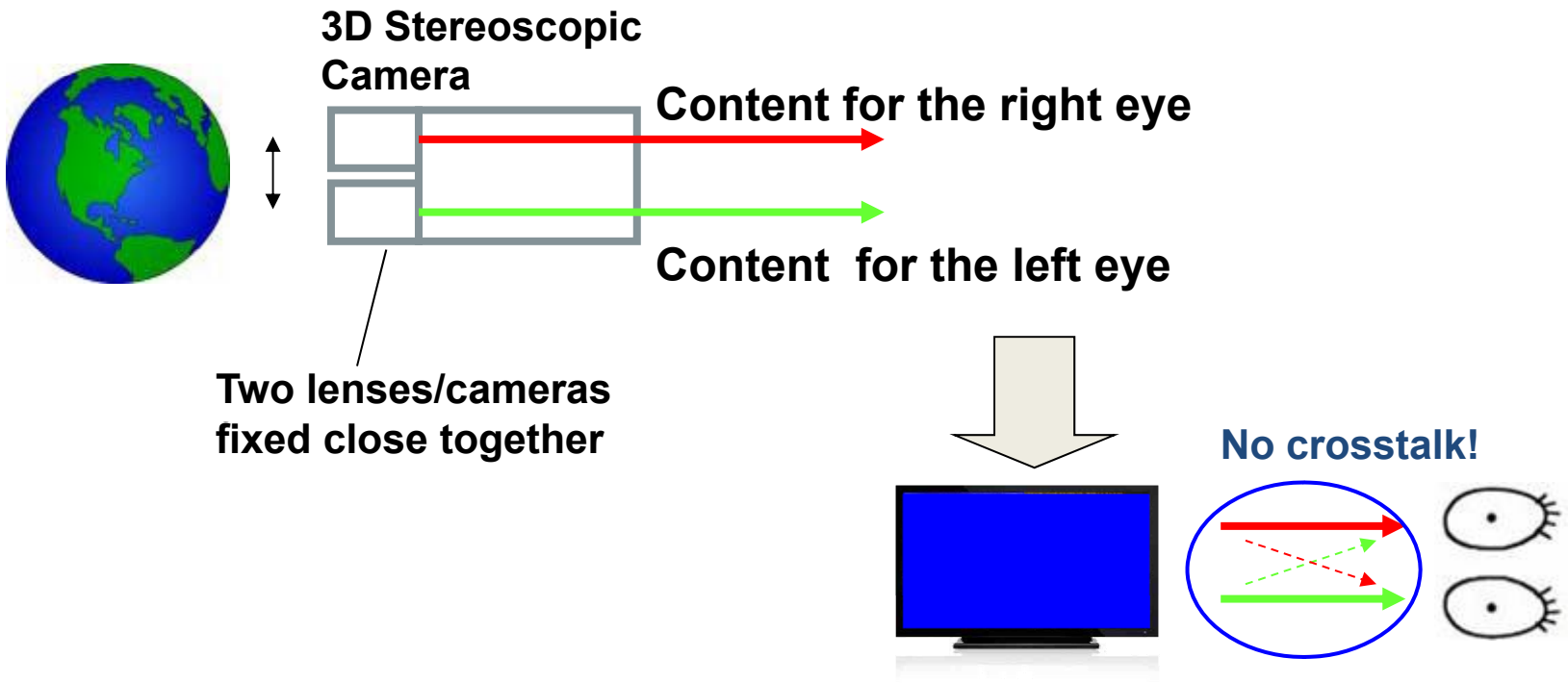
Stereo Camera with Converging Axis



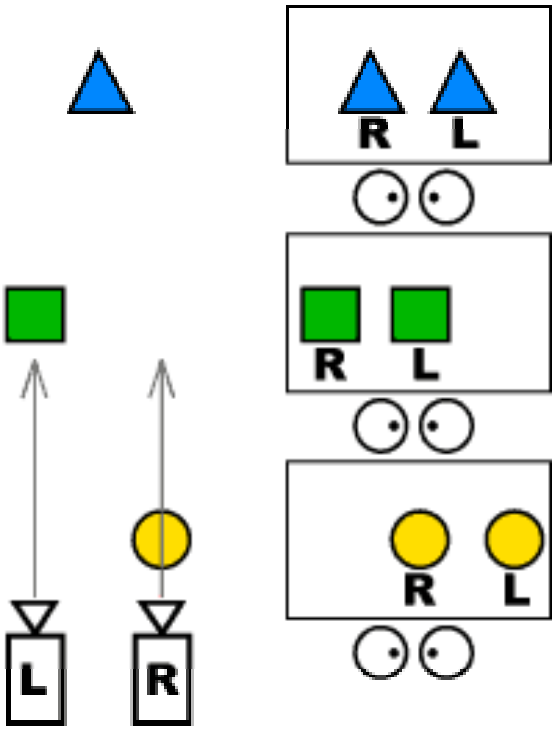
- **Keystone-Effect**
  - Requires undistortion

Principles of 3D: Stereoscopic Filming

- **3D is based on stereoscopic content**
  - 2 views recorded from each scene:

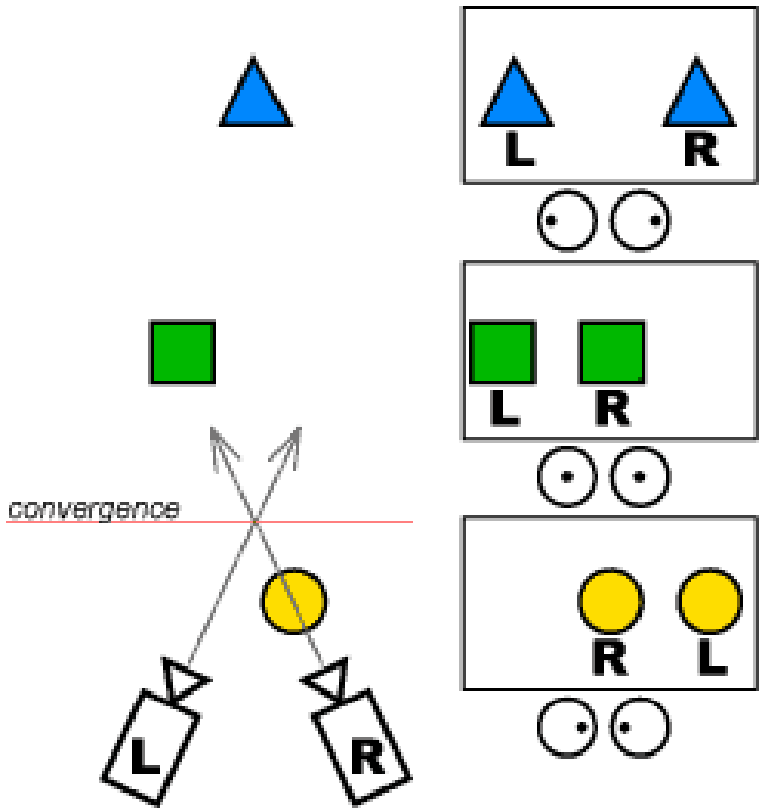


3D Parallel Filming



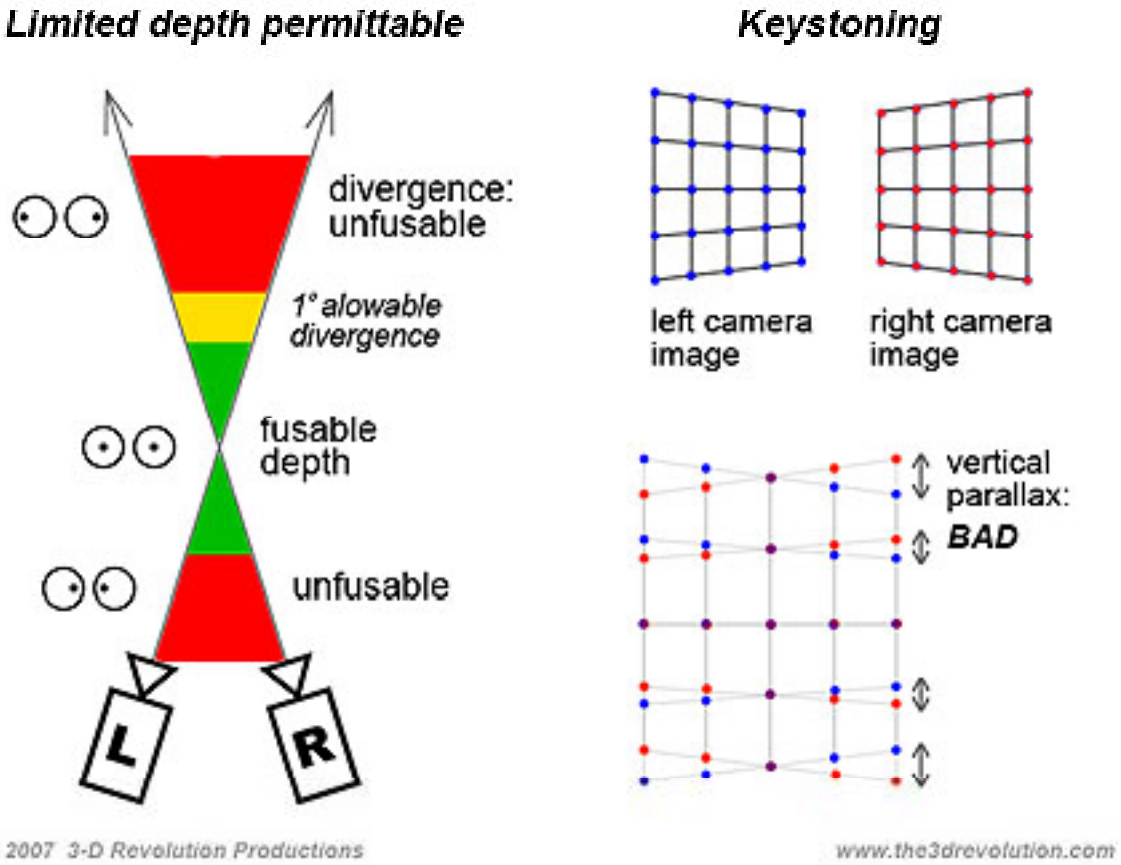
- **One Camera per eye**
  - ~63mm apart
- **Minimal Distortions**
- **Fixed viewing focal point**
- **Easier to view**
- **Can Be Less Realistic**

3D Converging Filming



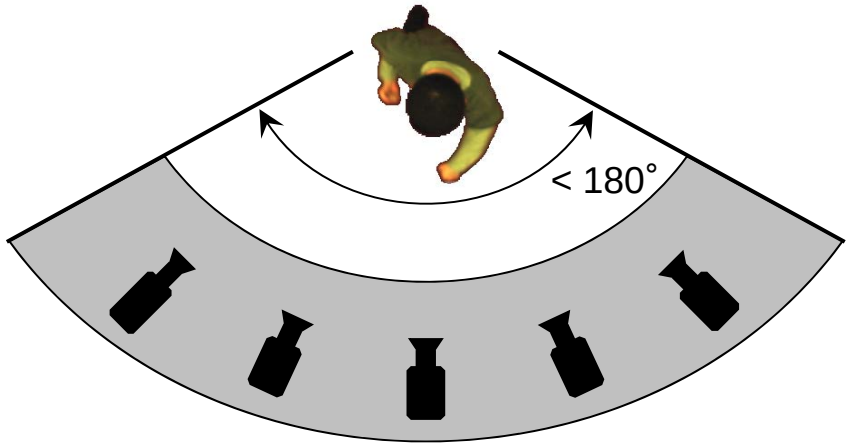
- **One Camera per eye**
  - ~63mm apart
- **Keystoning Distortions**
- **Variable Focus Point**
- **Can provide best picture**
- **But eye strain for near/far objects due to divergence**

3D Convergence Filming

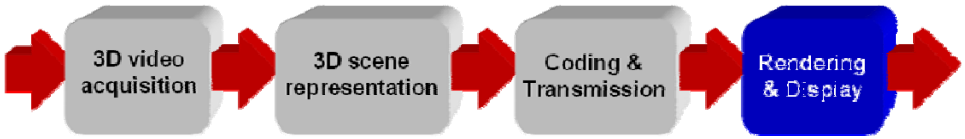


- **Provides Potentially Best Picture Quality**
- **Keystoning distortion may need post production correction**
- **Focussing on Near and Far objects may cause eye strain**

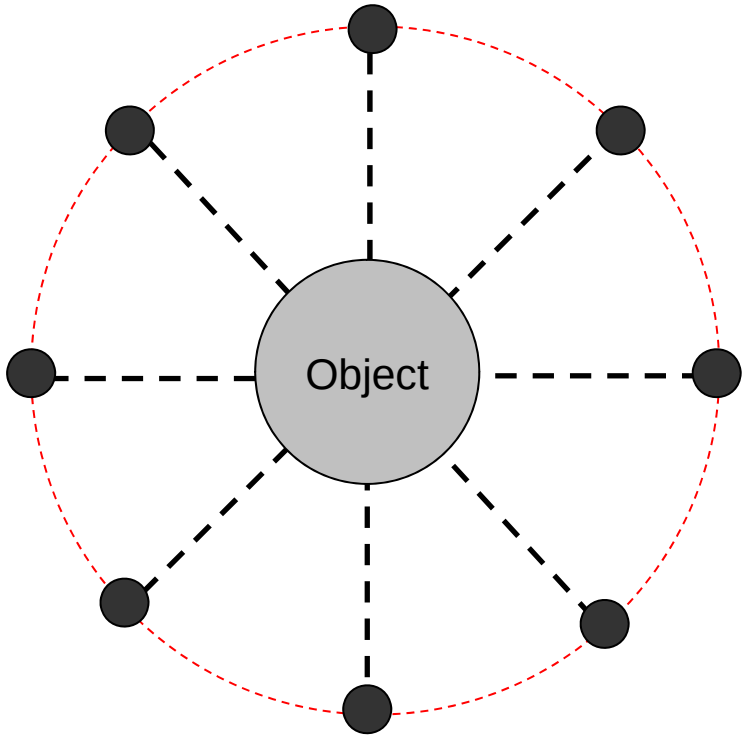
Free-View 3D Capture



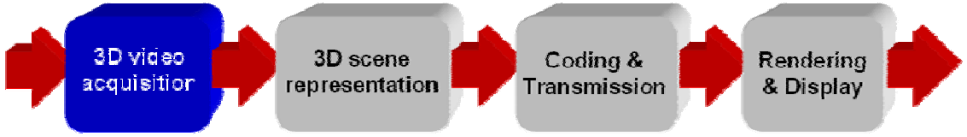
Viewpoint selection



- 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

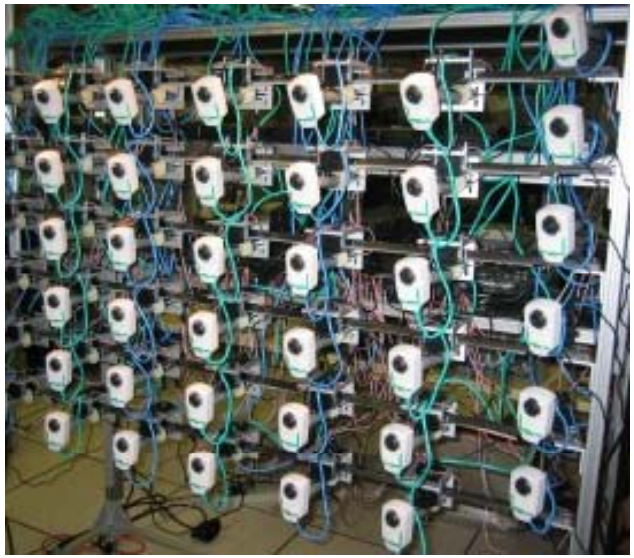


Stereo cameras



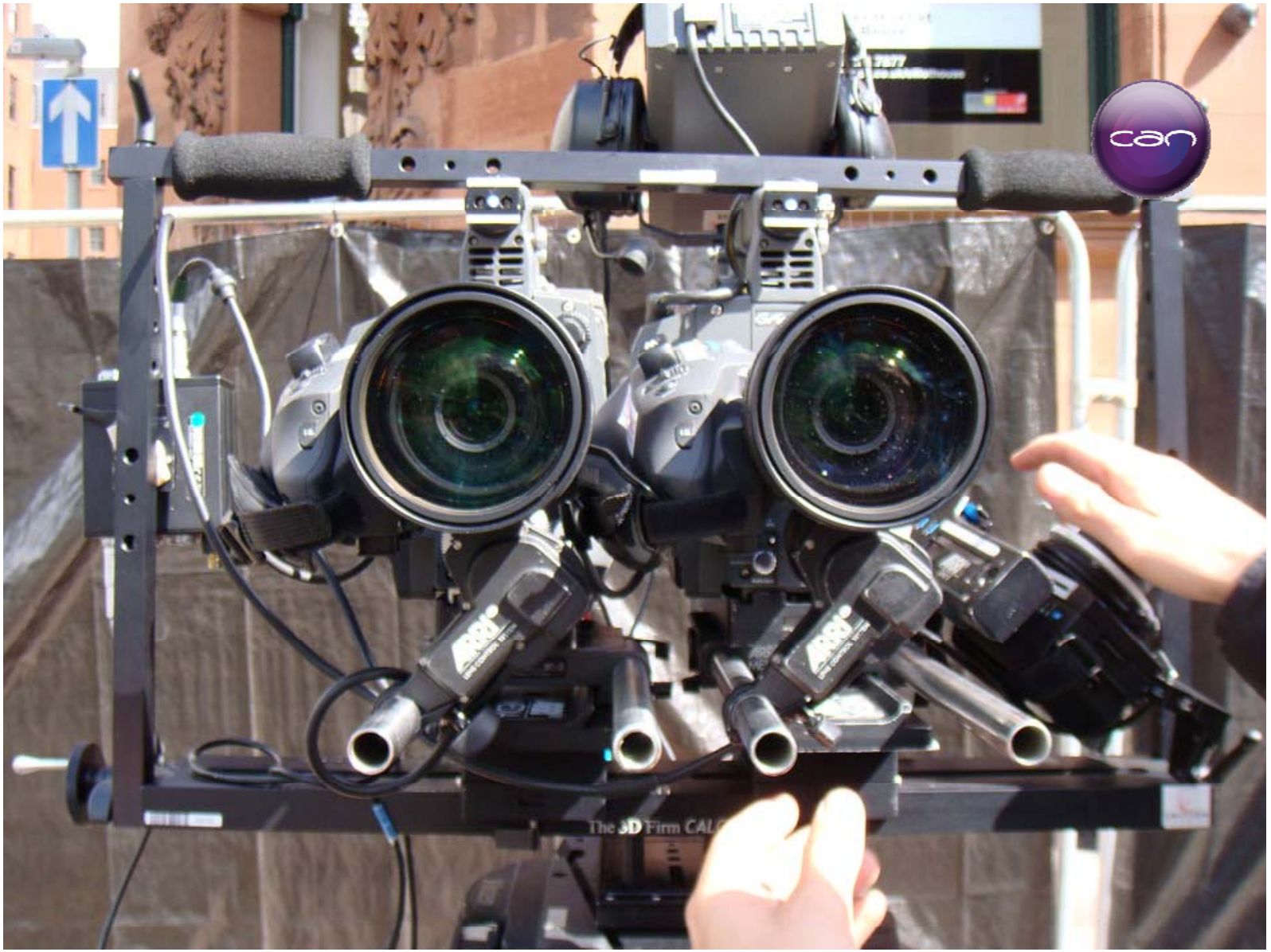
IMAX 3D camera

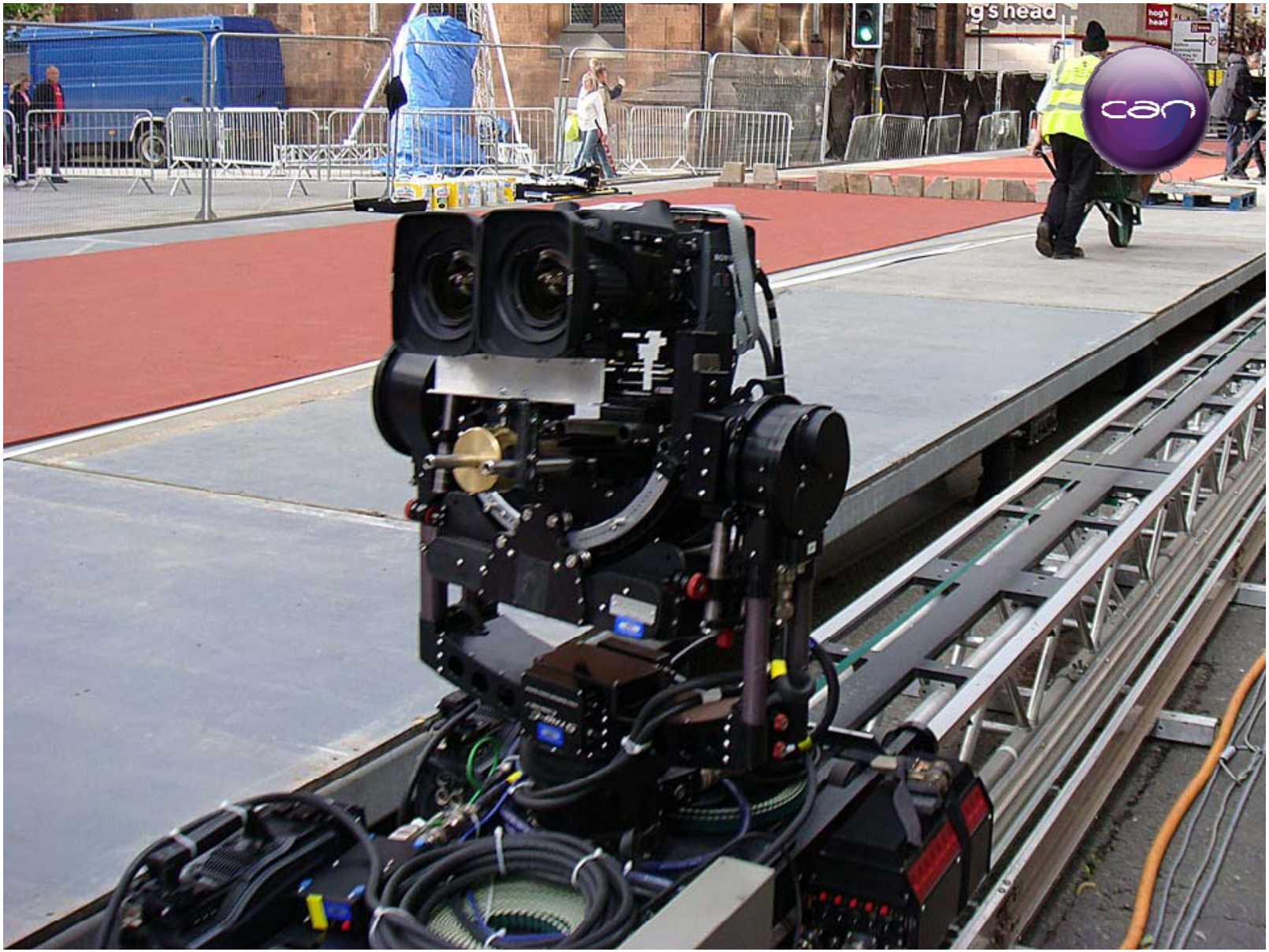
NASA Mars rover



Multiview camera array (Carnegy Mellon)

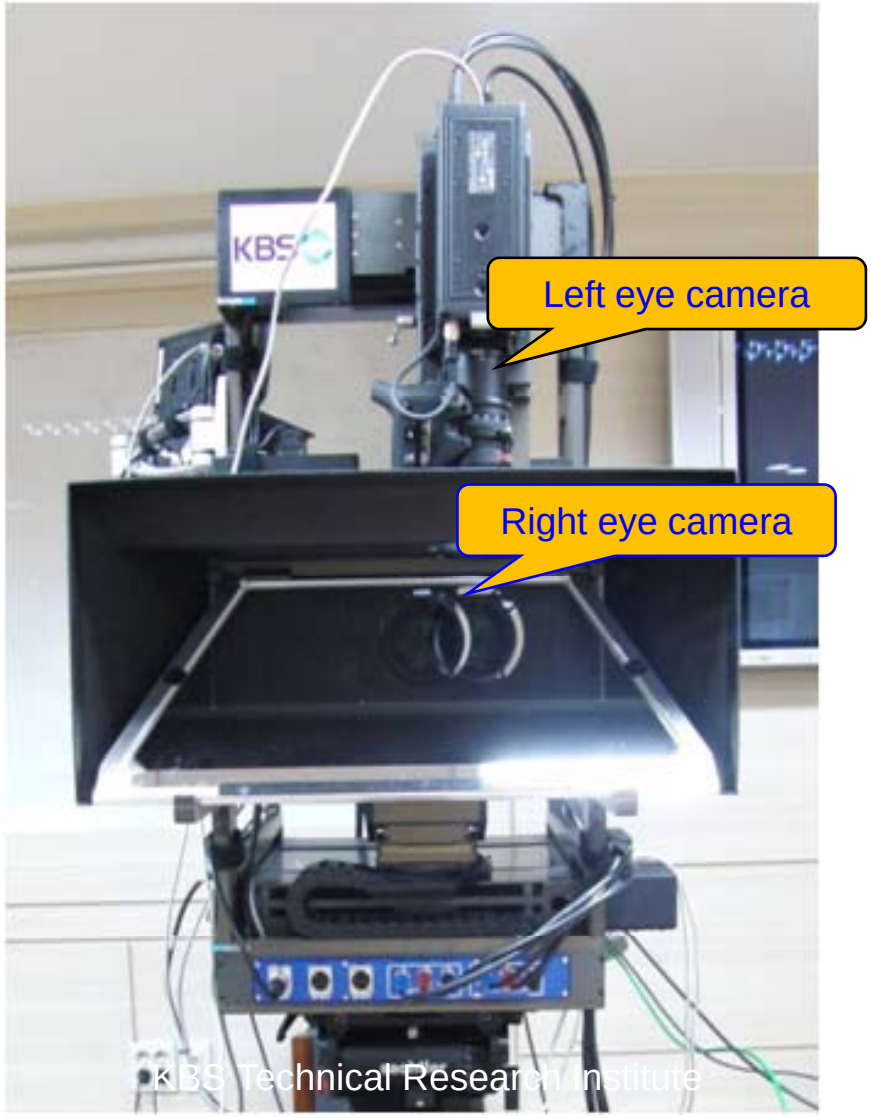
Stereo Camera




can

Stereo Camera

New 3DTV camera rig developed by KBS



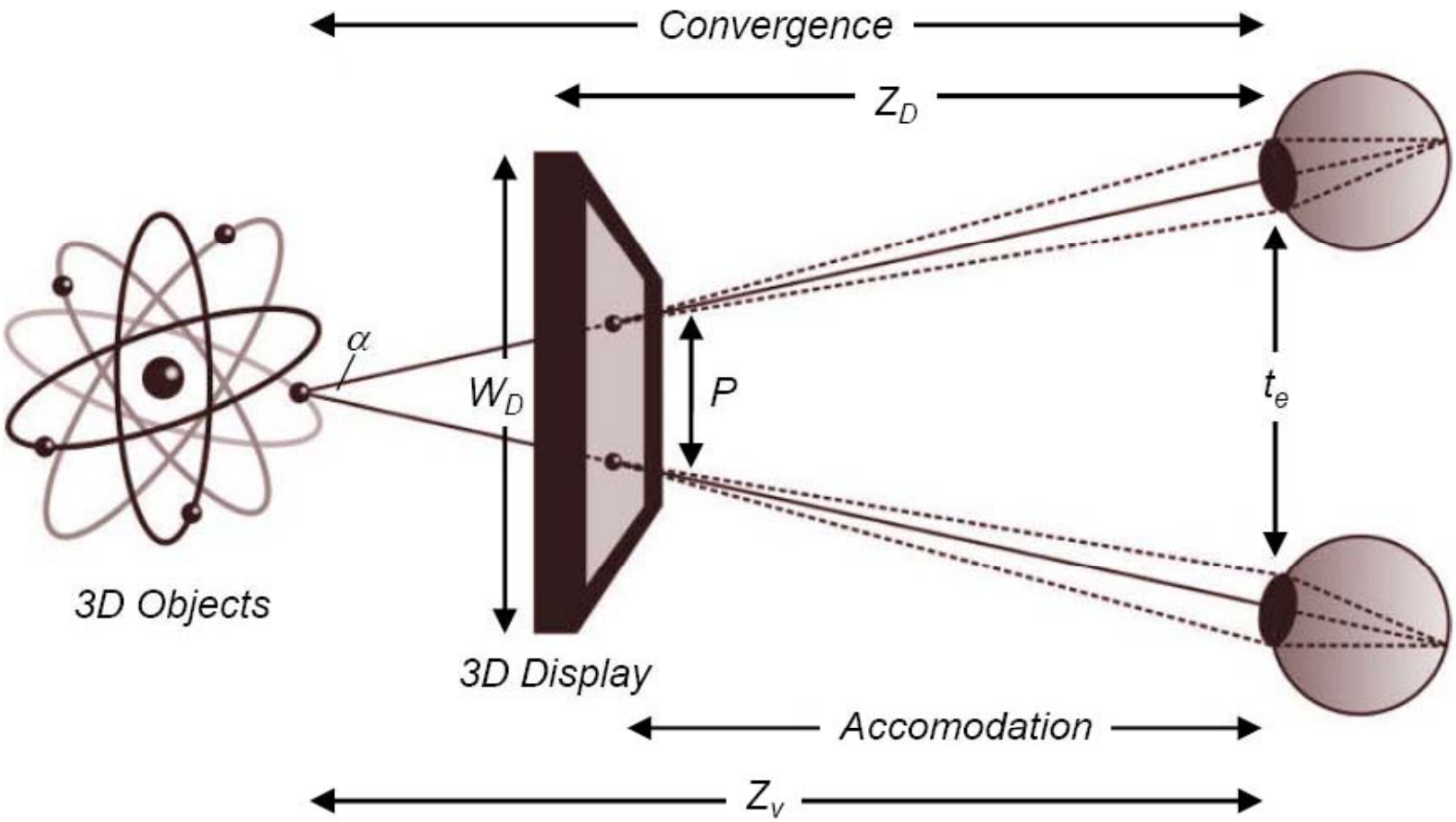
Signal

- Left and right image appear shifted
- Stereo enables humans to perceive shape

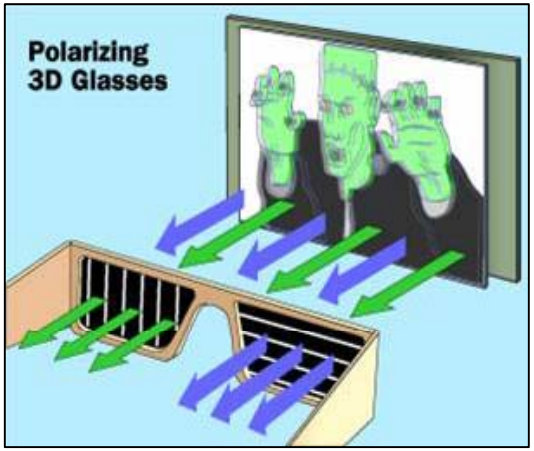
Contents

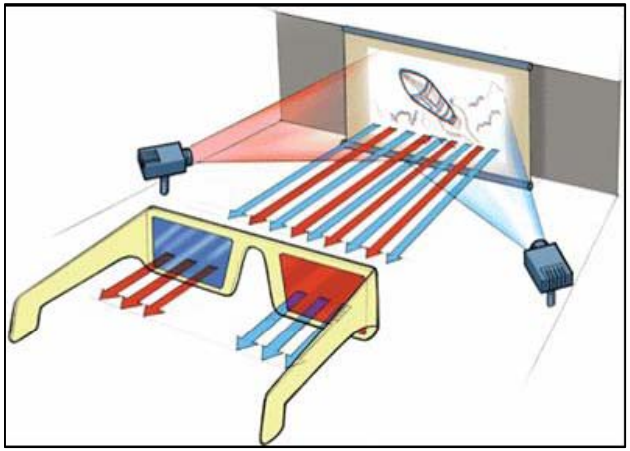
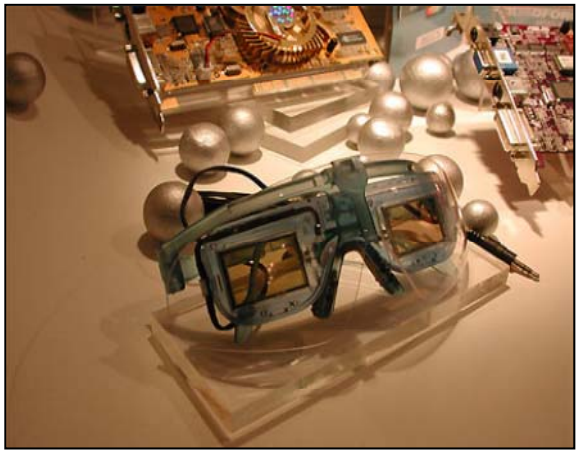
- Camera
- **Displays**
- Human Visual System
- Coding Concepts of Multiview Video
- 3DV
- Future Standards

The Real World on a 2D Screen



Eye-Glasses Stereoscopic



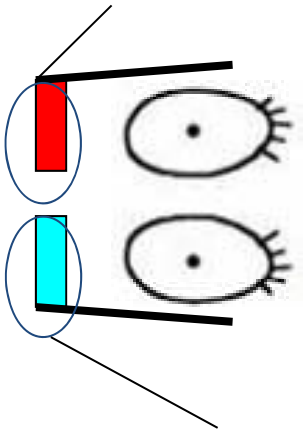
- Left: polarizing glasses
- Center: red-blue anaglyph
- Right: PC shutter-glasses

Passive 3DTV: Anaglyphic Glasses

Right eye content tinted red

Left eye content tinted cyan

Each Frame contains spatially interleaved left and right content



Red filter admits right content only

Cyan filter admits left content only

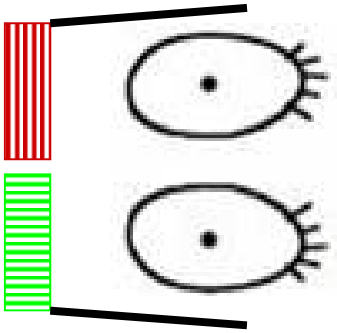
- **Advantages:**
  - Wide Viewing Angle
  - Lightweight and cheap glasses
  - Can be used with any type of Display
  - New complex colour filters provide excellent quality
- **Disadvantages:**
  - Poor Colour Fidelity with Red/Cyan Filters
  - More accurate/expensive displays for new complex filters

Passive 3DTV Displays: Polarizing Glasses

100/120Hz Frame Rate

Right Eye content
Clockwise
polarized by TV

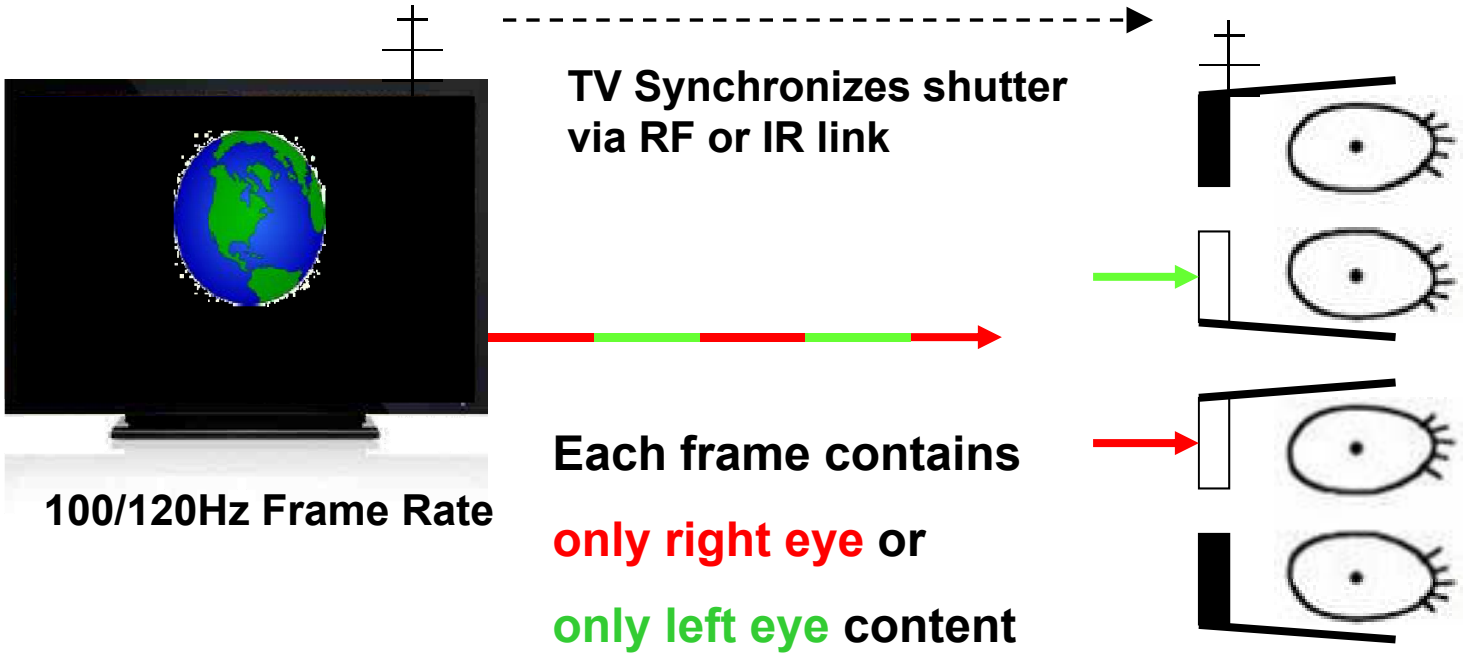
Left Eye content
Anticlockwise
polarized by TV



- **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)

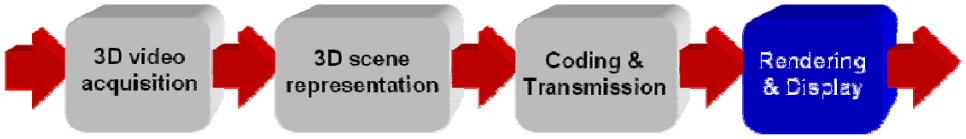
Active 3DTV Displays: Shutter Glasses

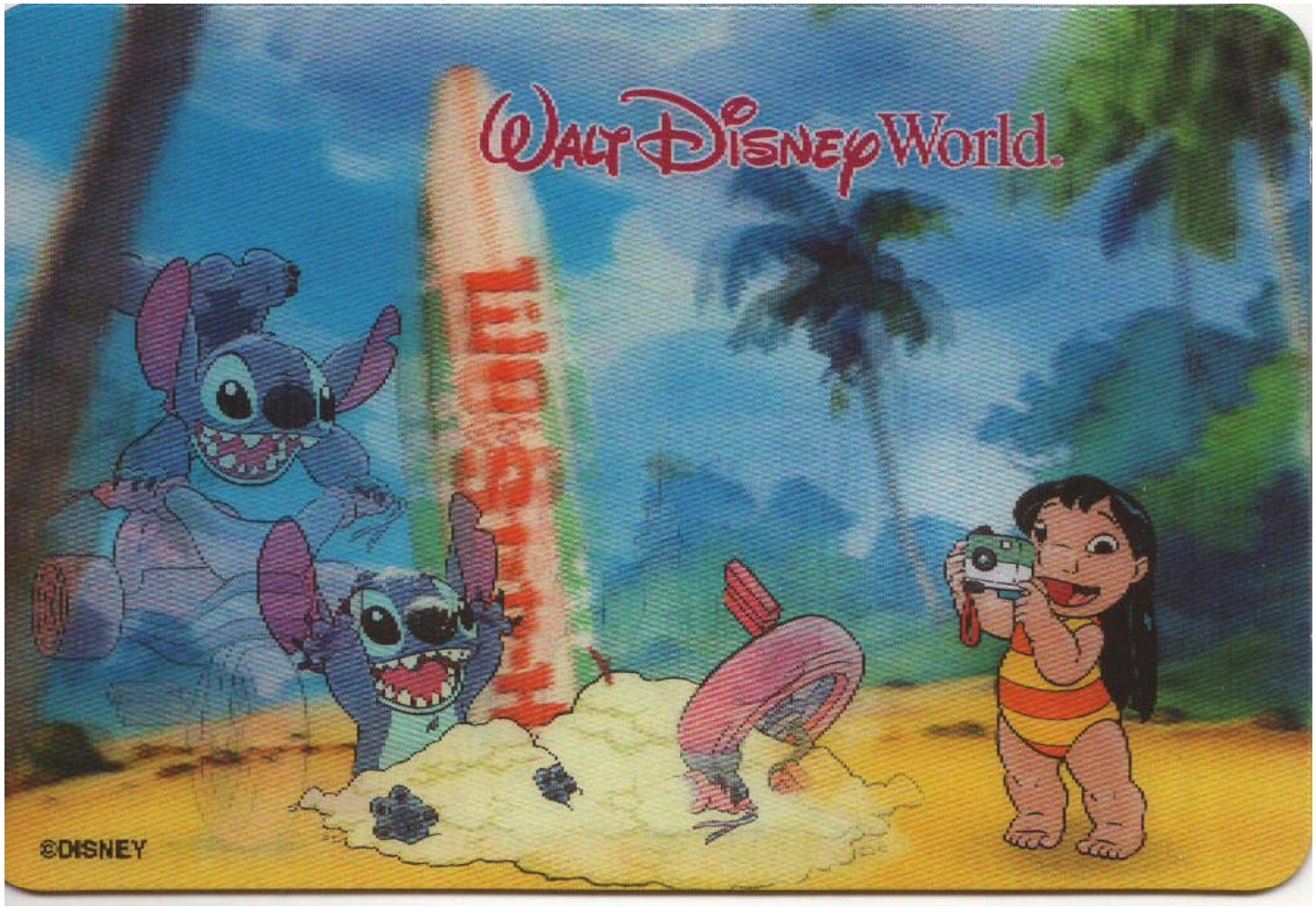


- **Advantages:**
  - Wide Viewing Angle
  - High Picture Quality on Any Display

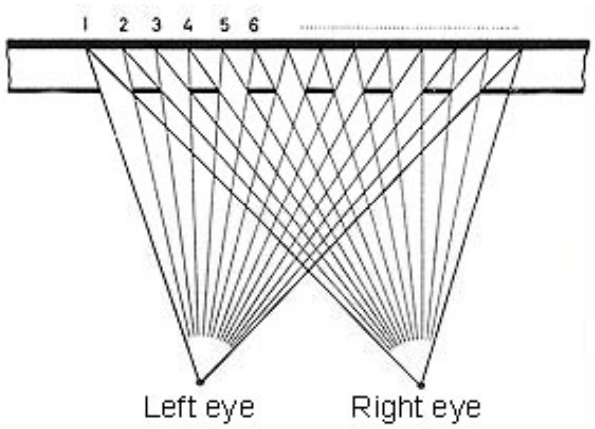
- **Disadvantages:**
  - LCD shutter response time can limit frame rates
  - Expensive battery powered glasses

Do we need glasses?

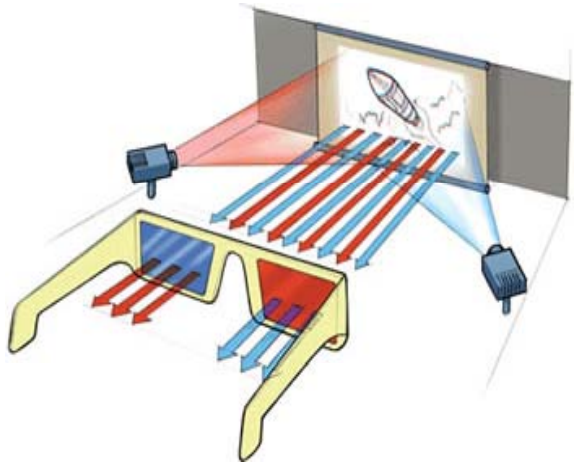




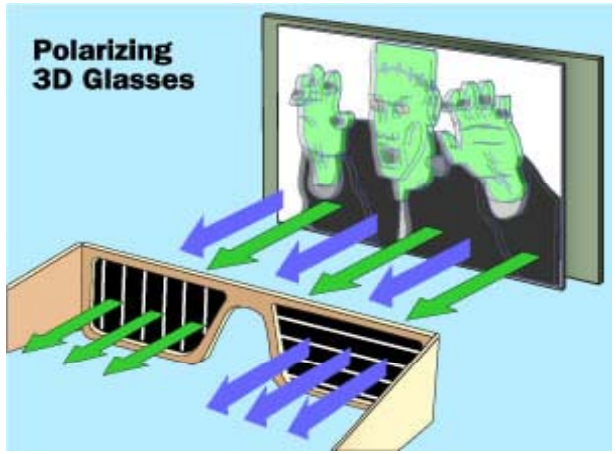
1950s: 3D Projection Systems



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

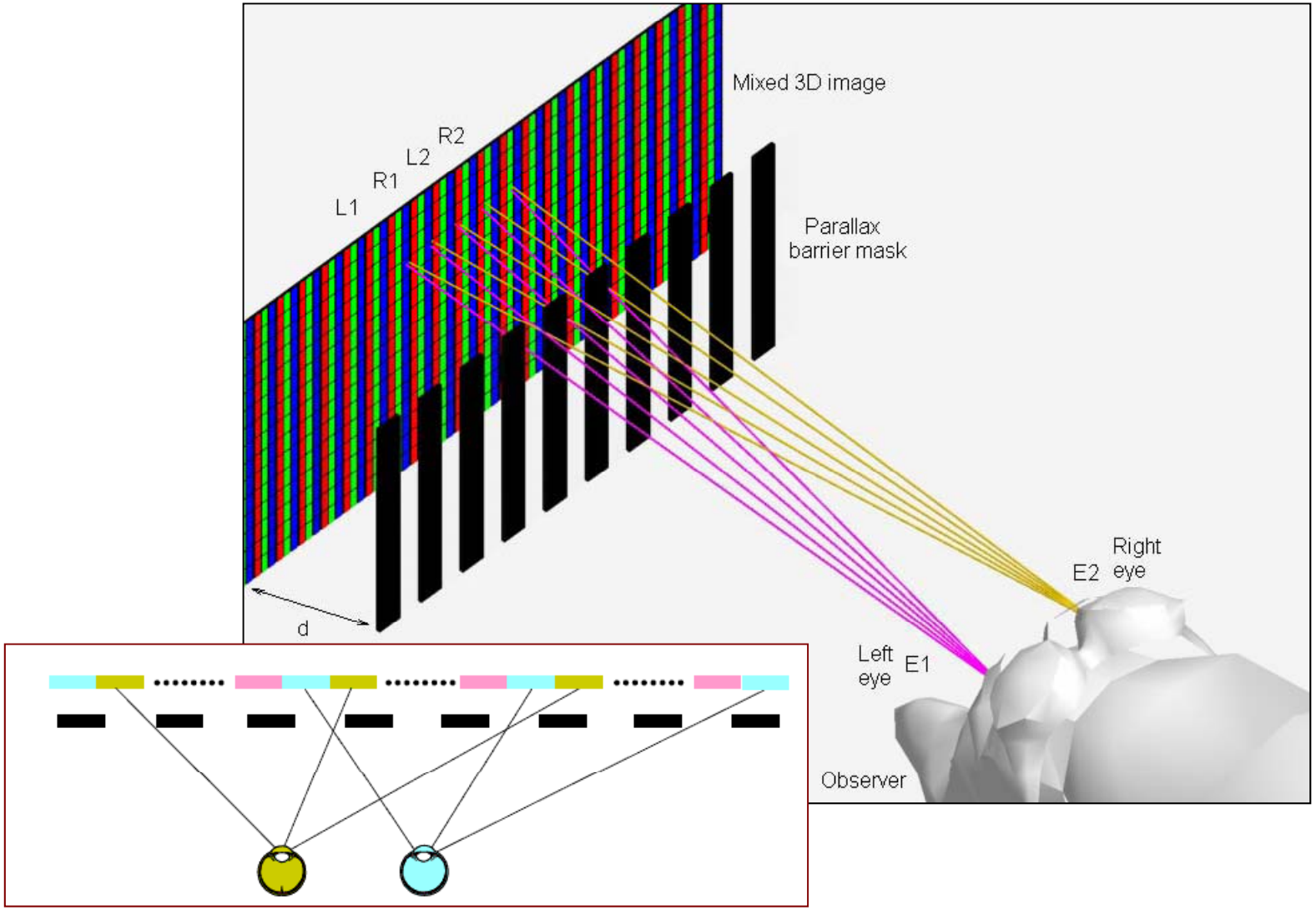


Anaglyphic Glasses

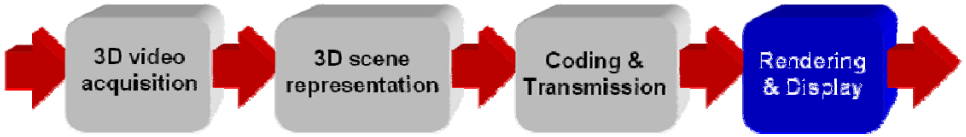


Polarizing Glasses

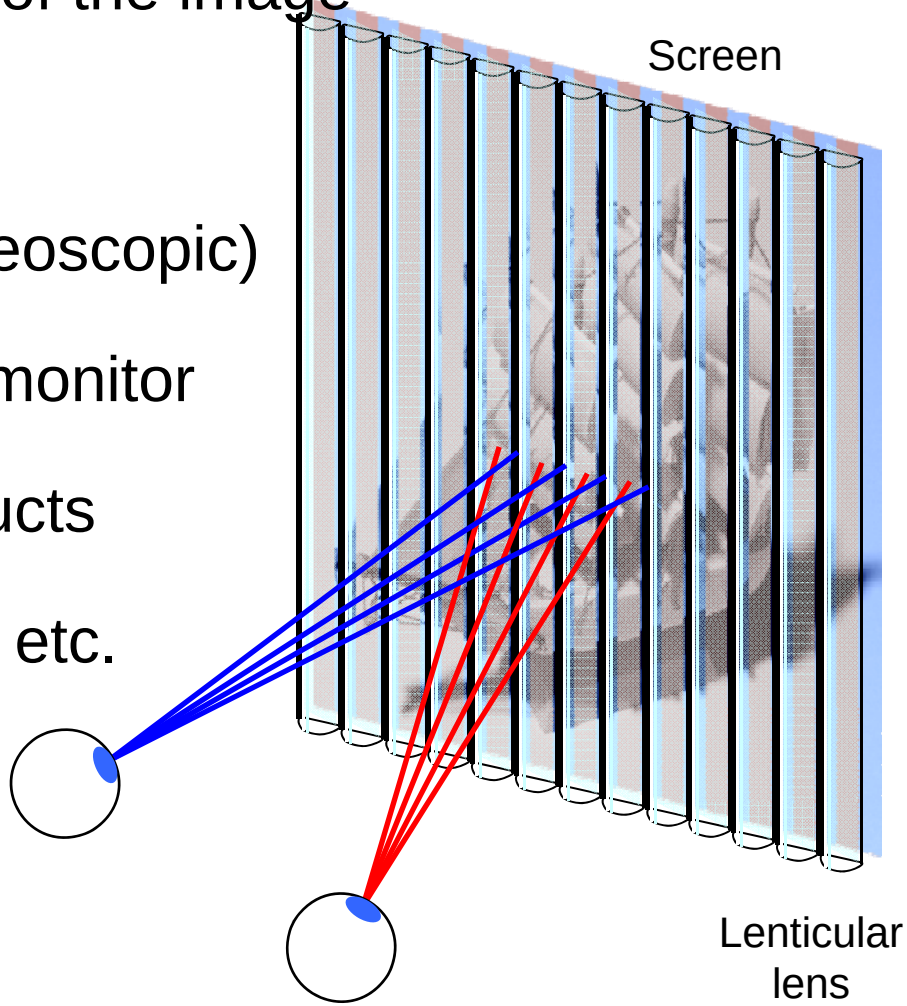
Parallax Barrier



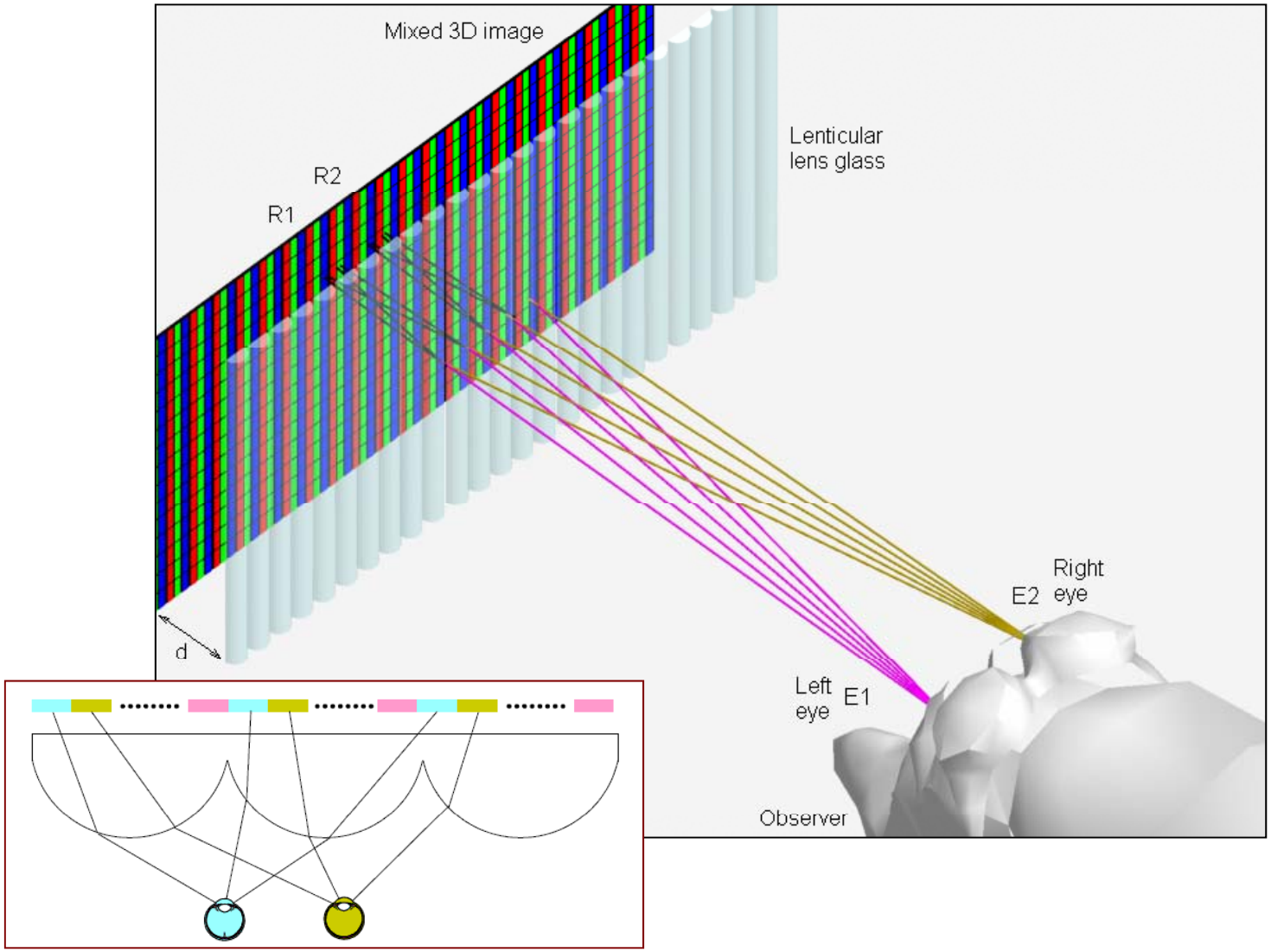
Lenticular display



- 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

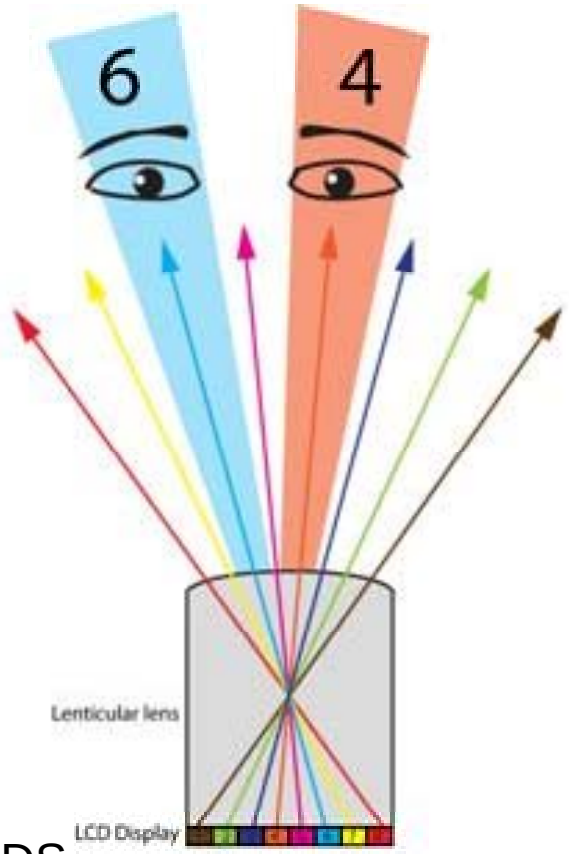


Lenticular


Mixed 3D image
Lenticular
lens glass
R2
R1
d
Right
E2 eye
Left
eye
E1
Observer

3DTV: Autostereoscopic Displays

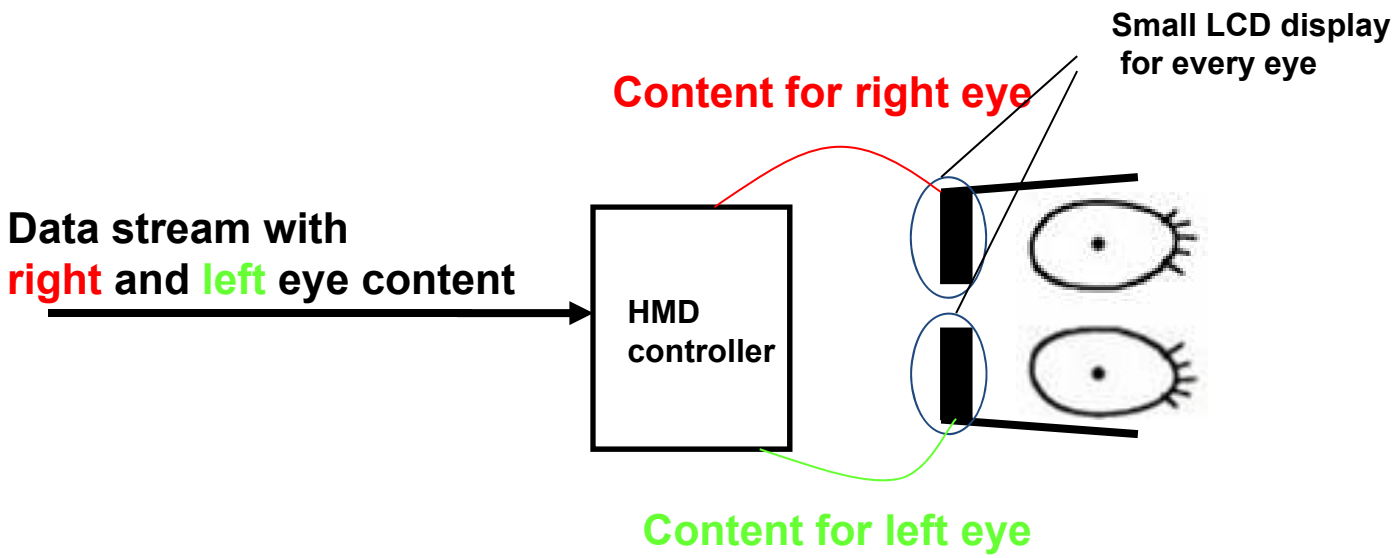
- Array of lenses fixed to display
  - Large CRT, Plasma and LCD
- Advantages:
  - No glasses required
  - Multiple Views possible
  - Viewer Tracking possible
- Disadvantages:
  - Reduced Display Resolution
  - Viewing Angle very Critical
  - Cross talk
- October 2009: New 3D OLEDs from 3M
  - Wider 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

3DTV: 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 3DTV review
http://gizmodo.com/5501900/the-best-3dtv-samsung-un55c7000-vs-panasonic-tc+p50vt20

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

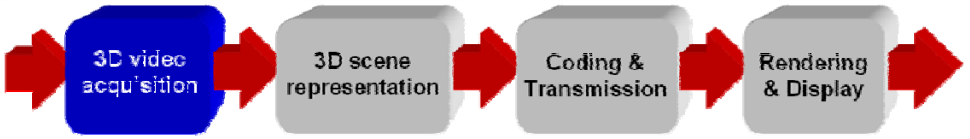
➢ Autostereoscopic

- Display-Type
  - Lenticular
  - Parallax Barrier
  - Integral Photography (IP)
  - Grating Array Plates
- Projection-Type
  - Fresnel Lenz
  - Holographic Screen
  - Spherical Mirror

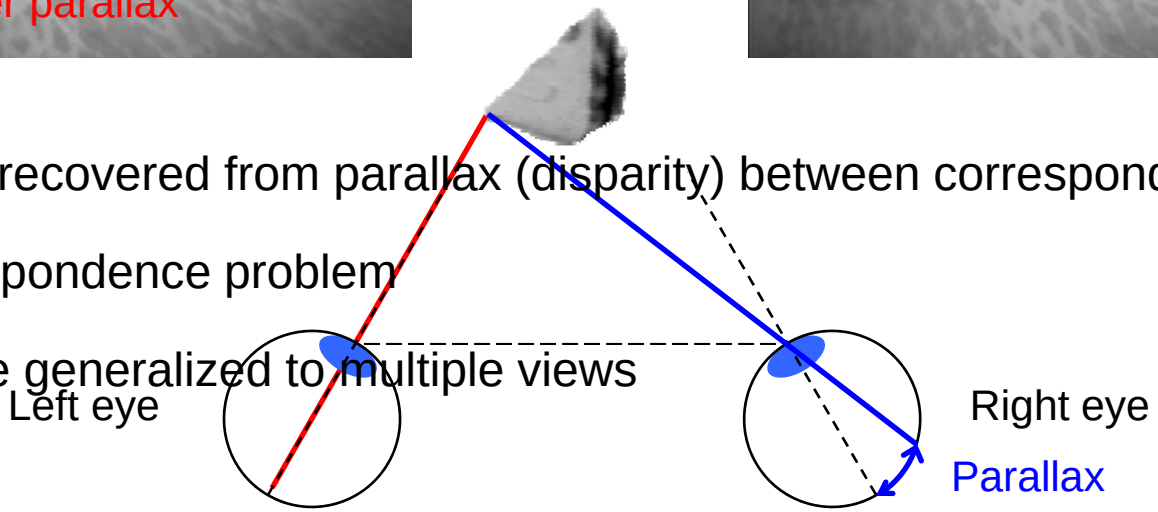
Contents

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

Shape from stereo

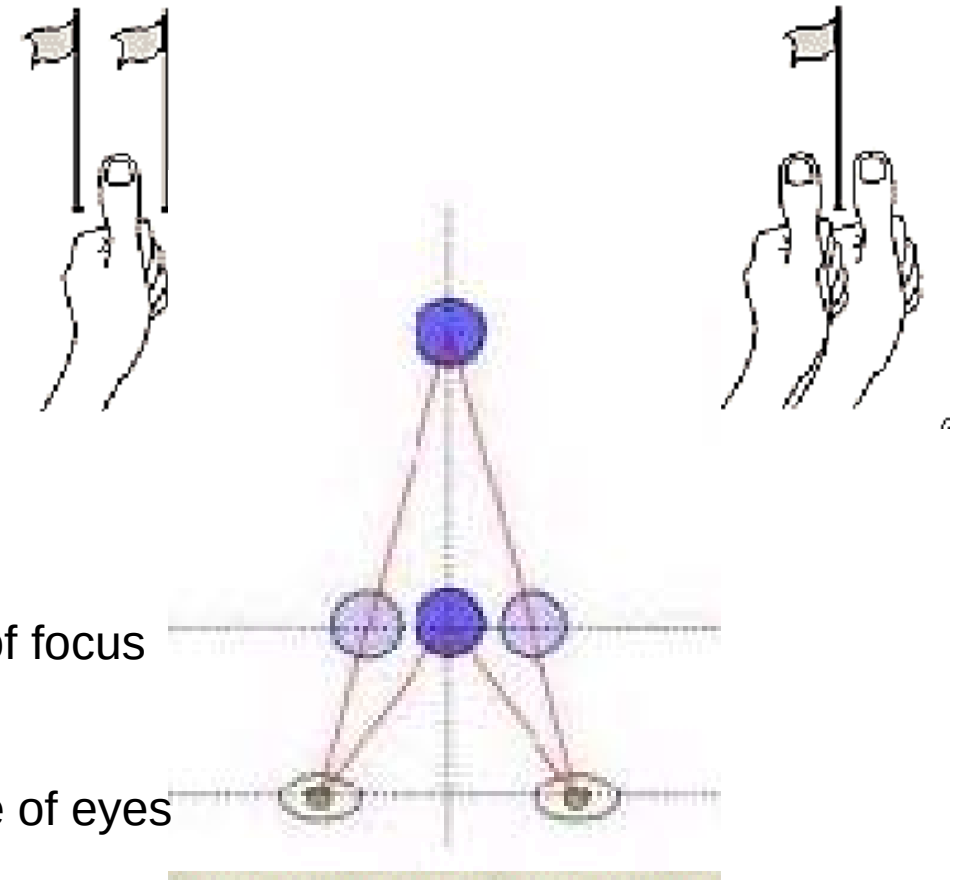


- Depth recovered from parallax (disparity) between corresponding points
- Correspondence problem
- Can be generalized to multiple views

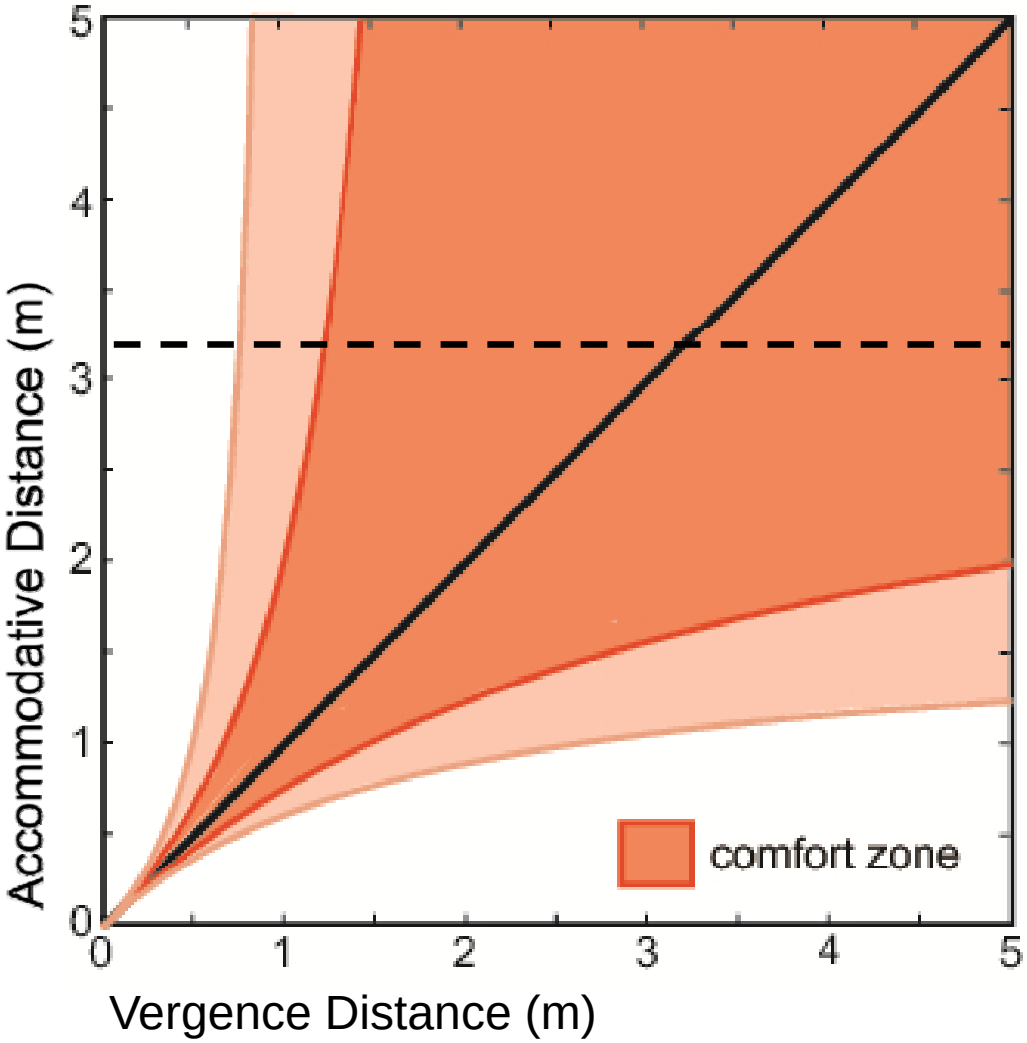


Depth Perception

- **Vergence: Eyes directed towards object**
- **Accomodation: Eyes focus the lens according to distance of object**
  - Limited depth of focus
  - Small sharp area

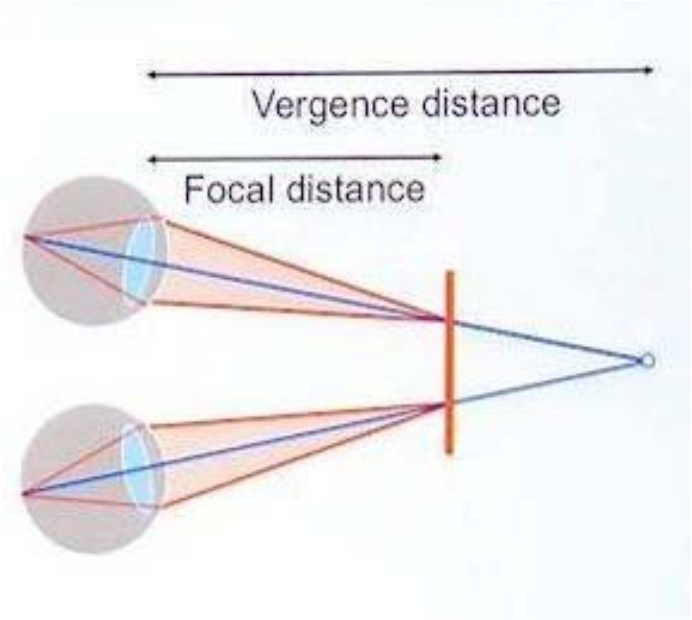


Vergence-Accommodation Conflict



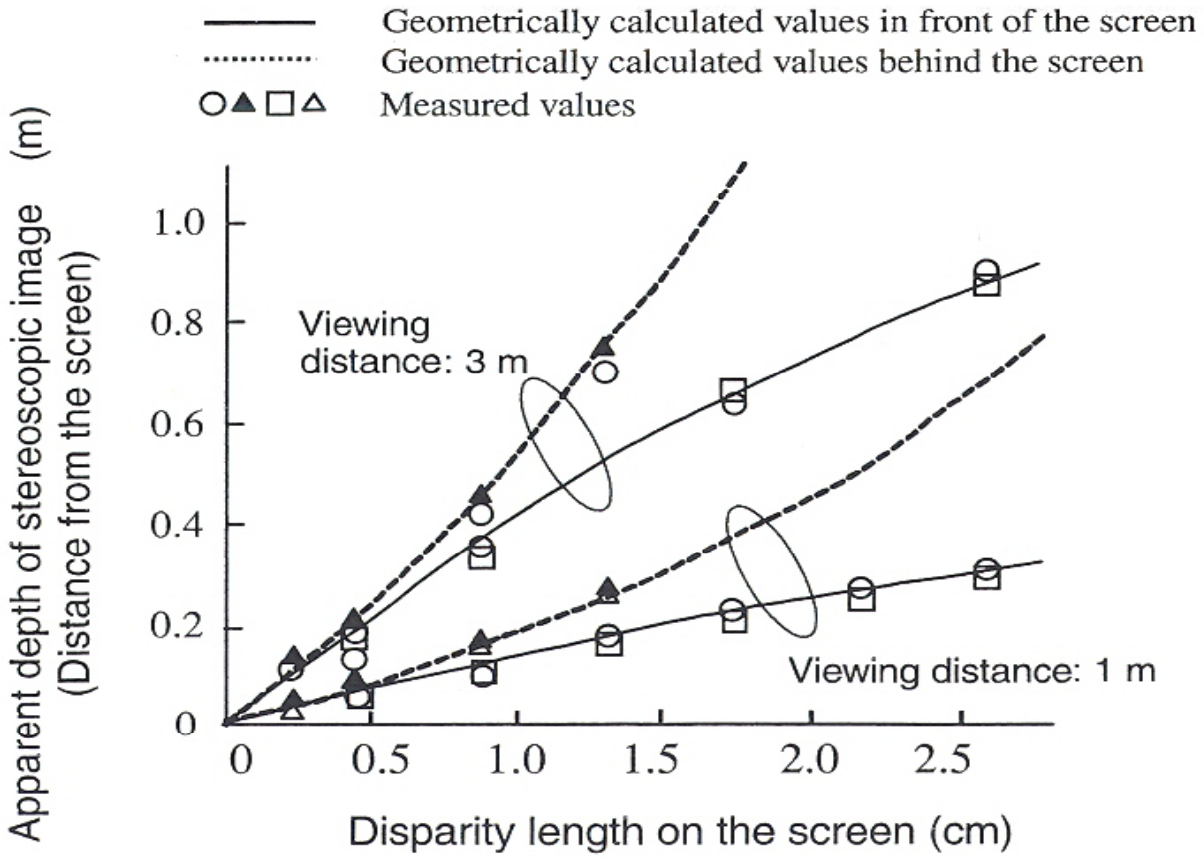
from Prof. Martin Banks,
Visual Space Perception Lab
University of California – Berkeley

3.2 meters (10.5 feet) ok to +∞

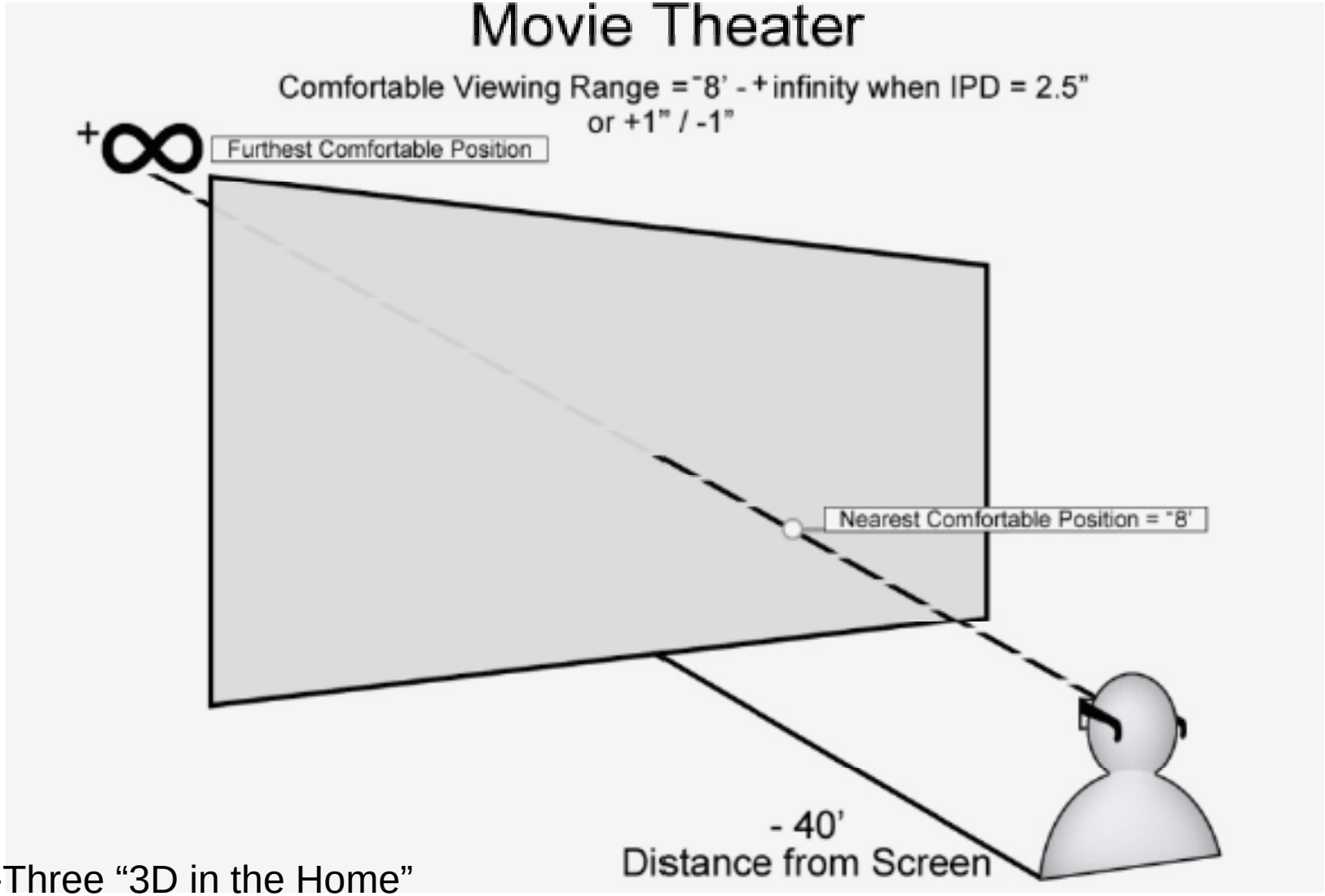


Why We Can See the Stereo TV?

- **The relation of the depth and the disparity length**

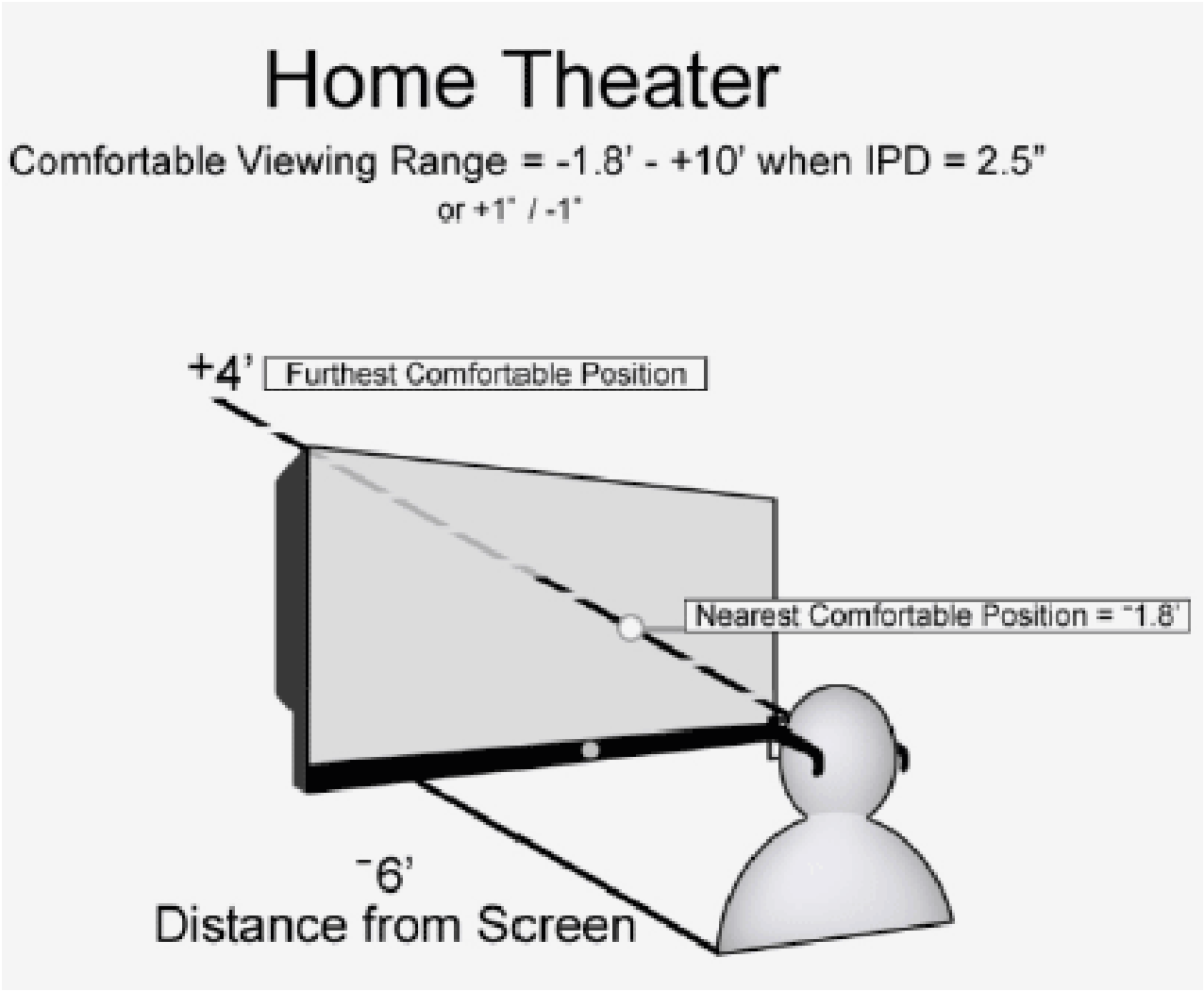


Countering V-A Conflict



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

Countering V-A Conflict

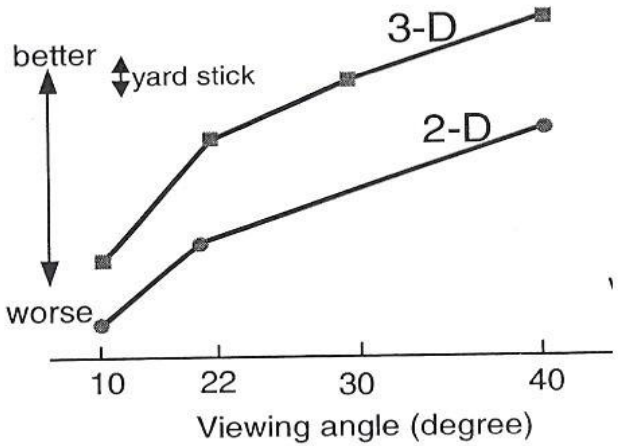


In-Three “3D in the Home”

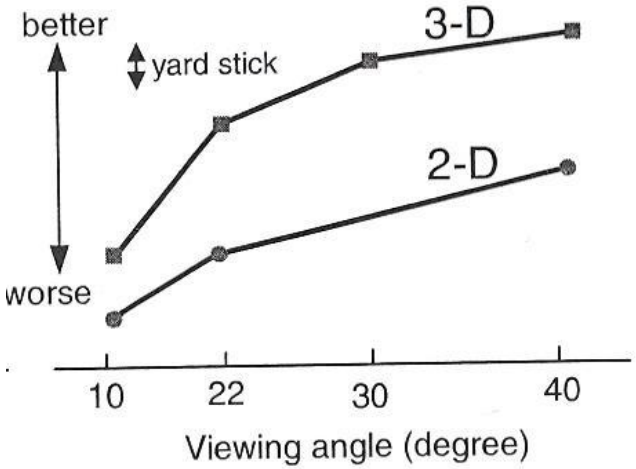
http://www.in-three.com/3DintheHomev2.html

Which One Is Better, 2D or 3D?

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



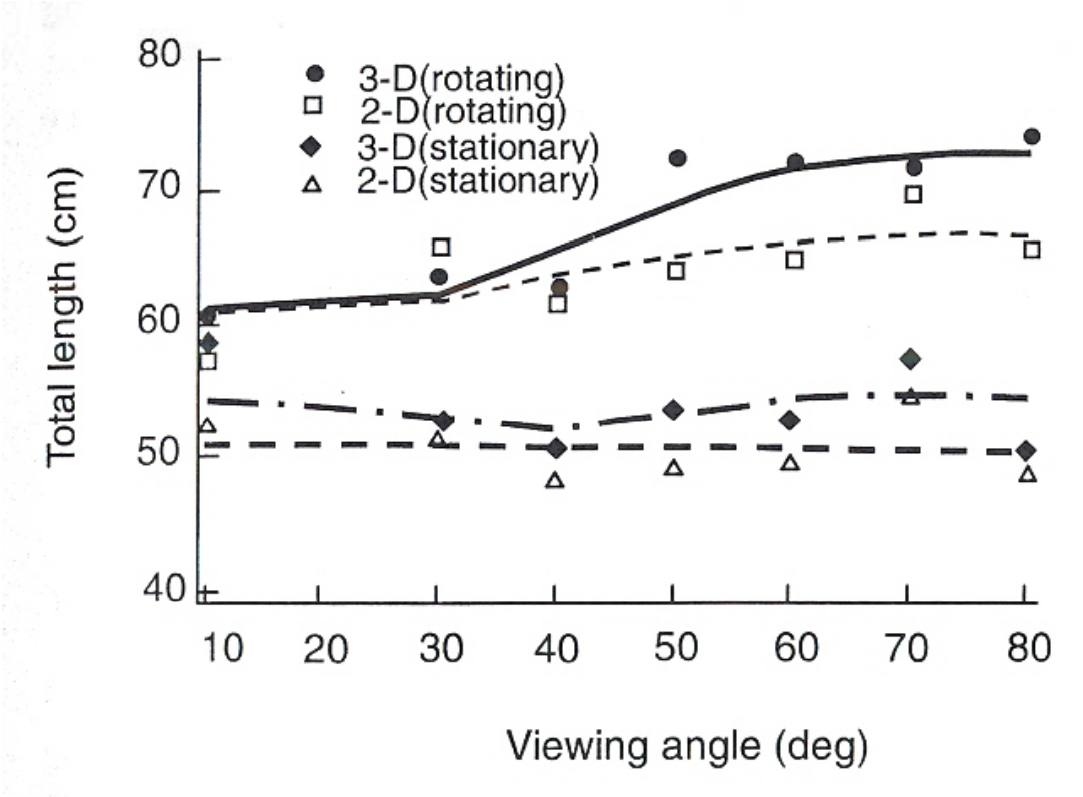
- Sensation of Power



- Total picture quality

Which One Is Better, 2D or 3D?

- Objective evaluation of Psychological Effect
- The body sway of the viewer is tracked, measured and analyzed.



Miniaturization

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

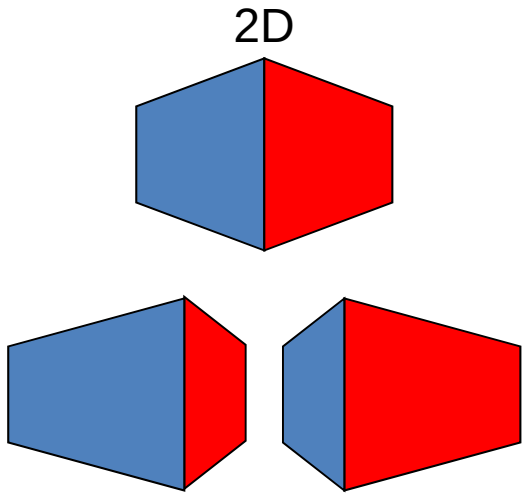


Image position relatively easy to change; image content much harder to change

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

Experimental Results (cont.)

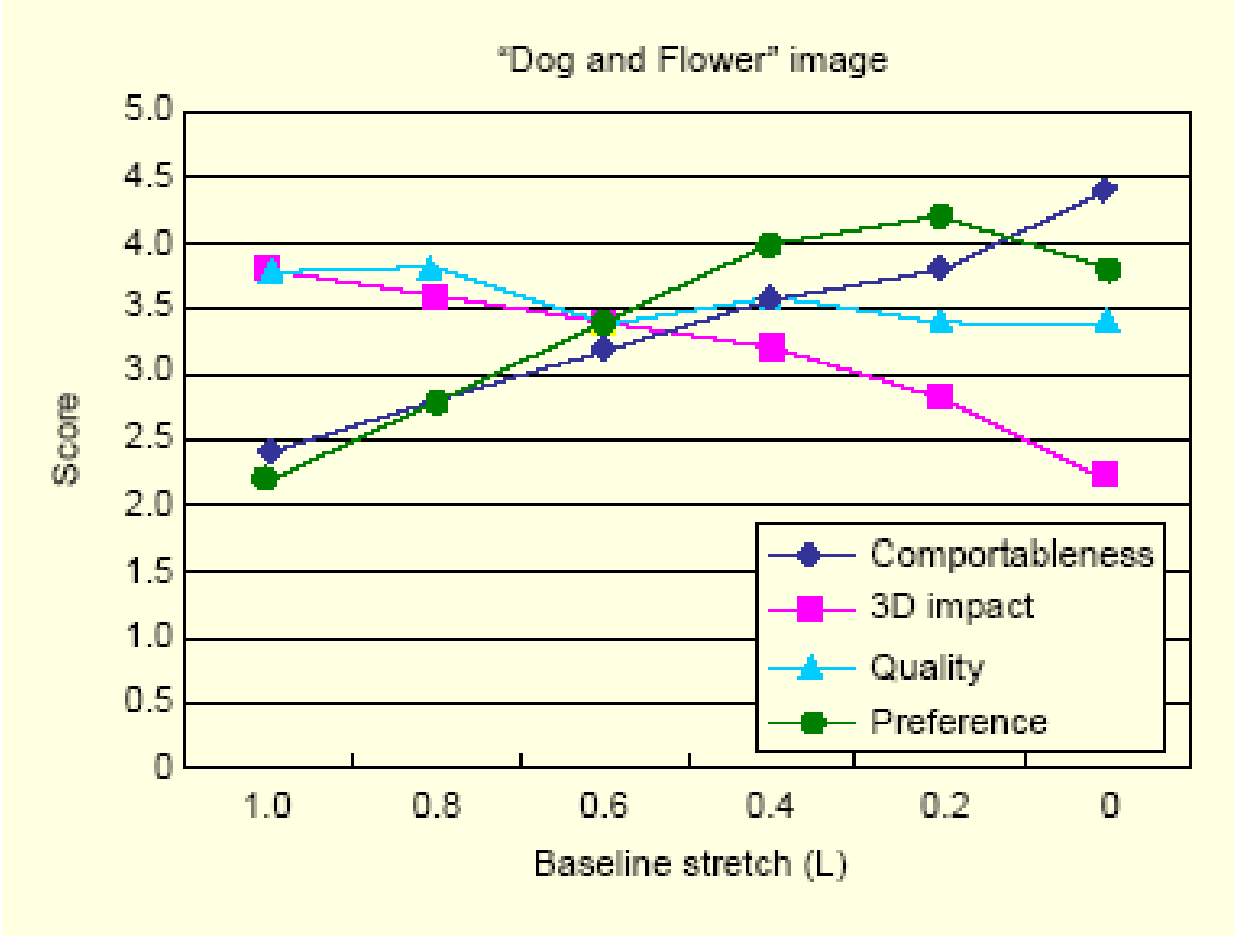


Chart 1. The average scores of subjective evaluation for the indoor stereoscopic scene

Experimental Results (cont.)

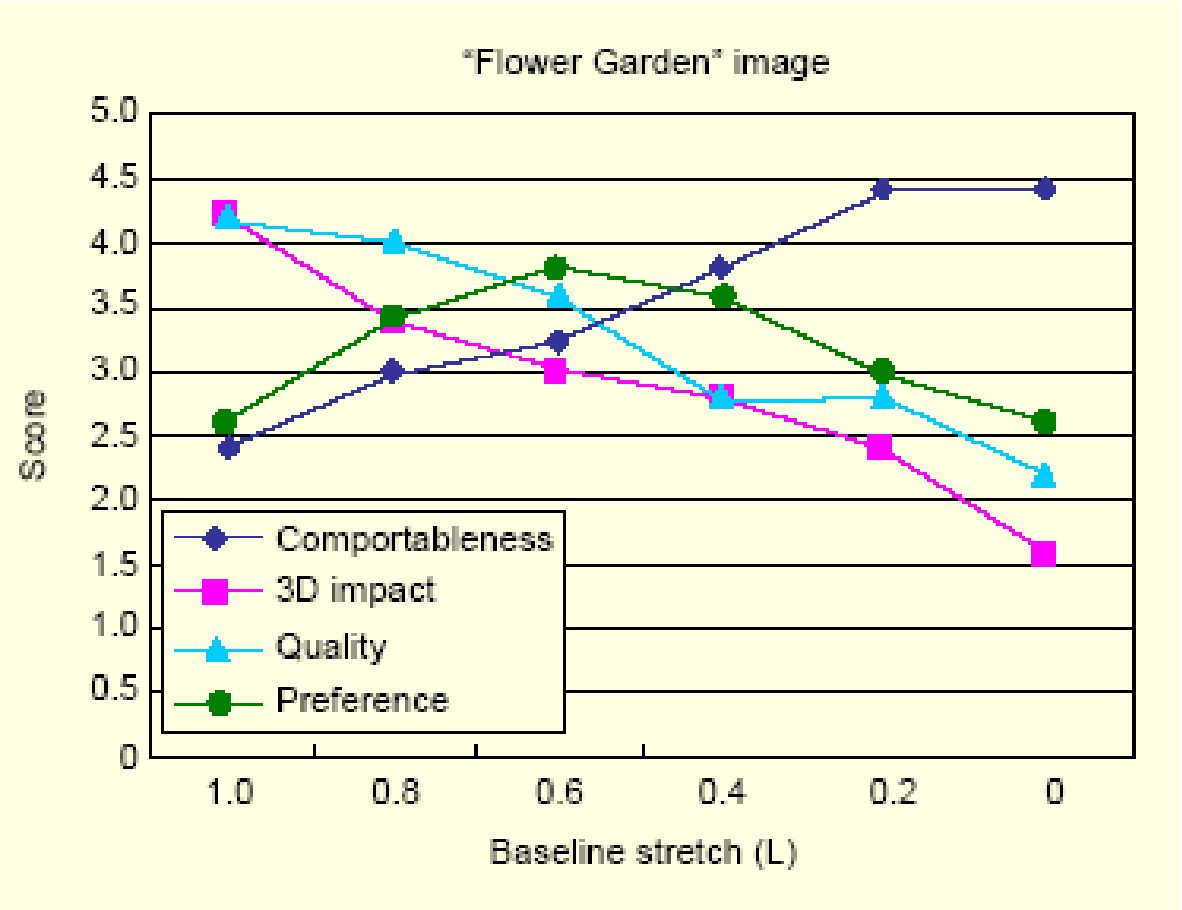
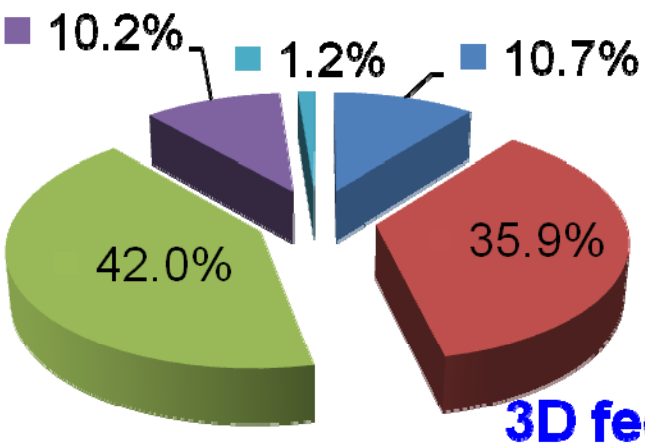
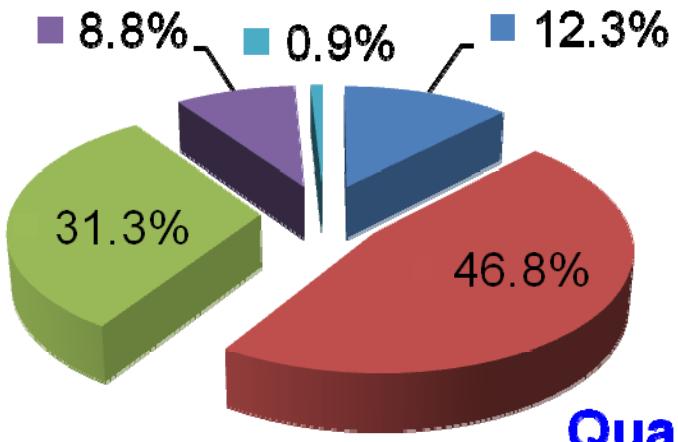


Chart 2. The average scores of our subjective evaluation for the outdoor HD stereoscopic scene

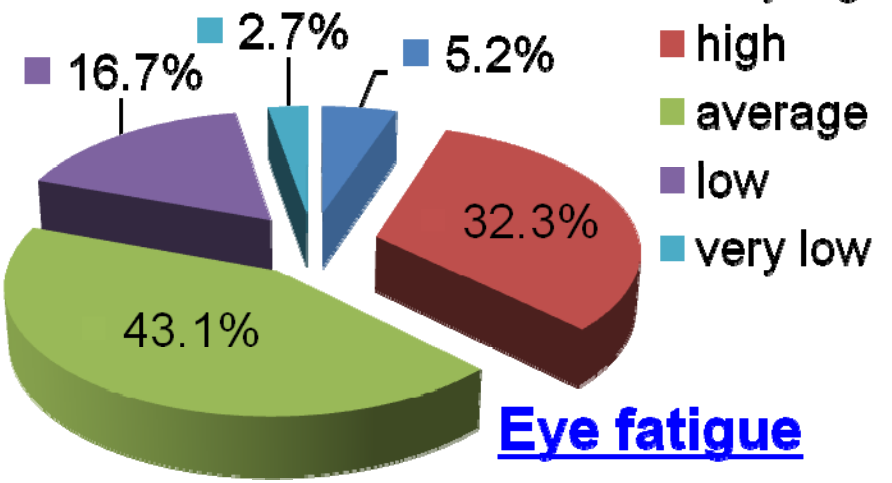
Subjective evaluation for 3DTV watching



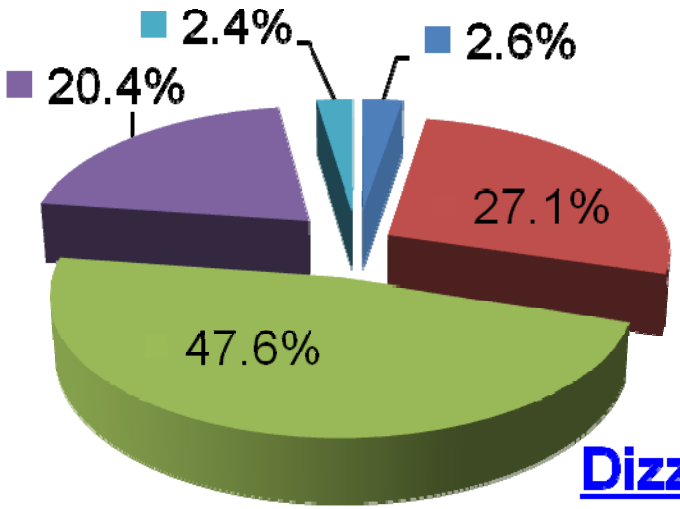
3D feeling



Quality



Eye fatigue



Dizziness

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

Eyes versus Stereo Camera

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

- 2 cameras
- No coupling
  - motors, interactivity, gaze tracking
- Disparity
- No
  - Motors, interactivity, gaze tracking

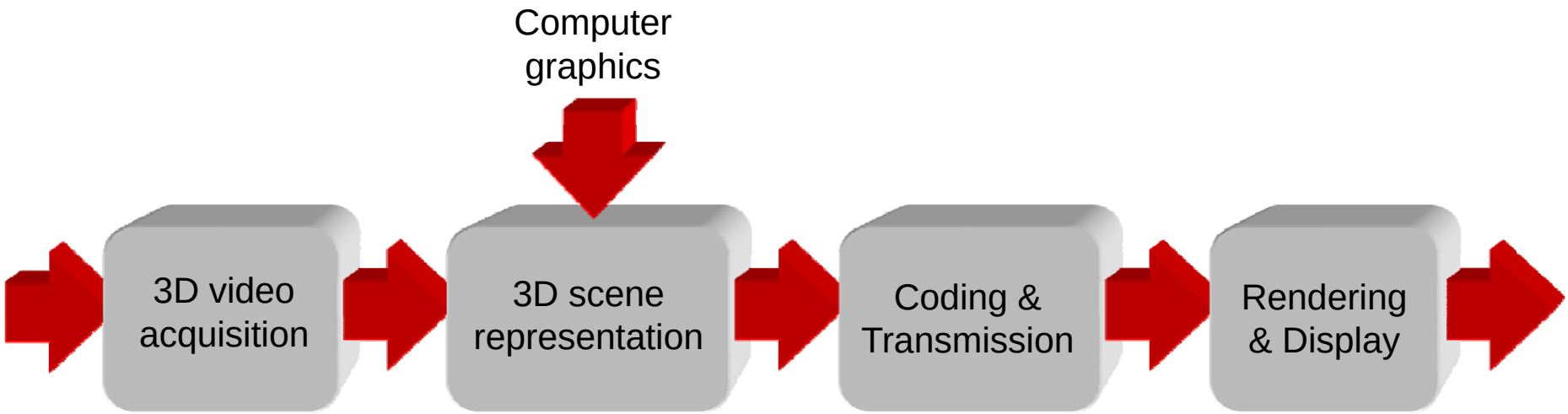
ITU-Generations (Profiles) of 3D-TV

- **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.

Contents

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

Coding & Transmission



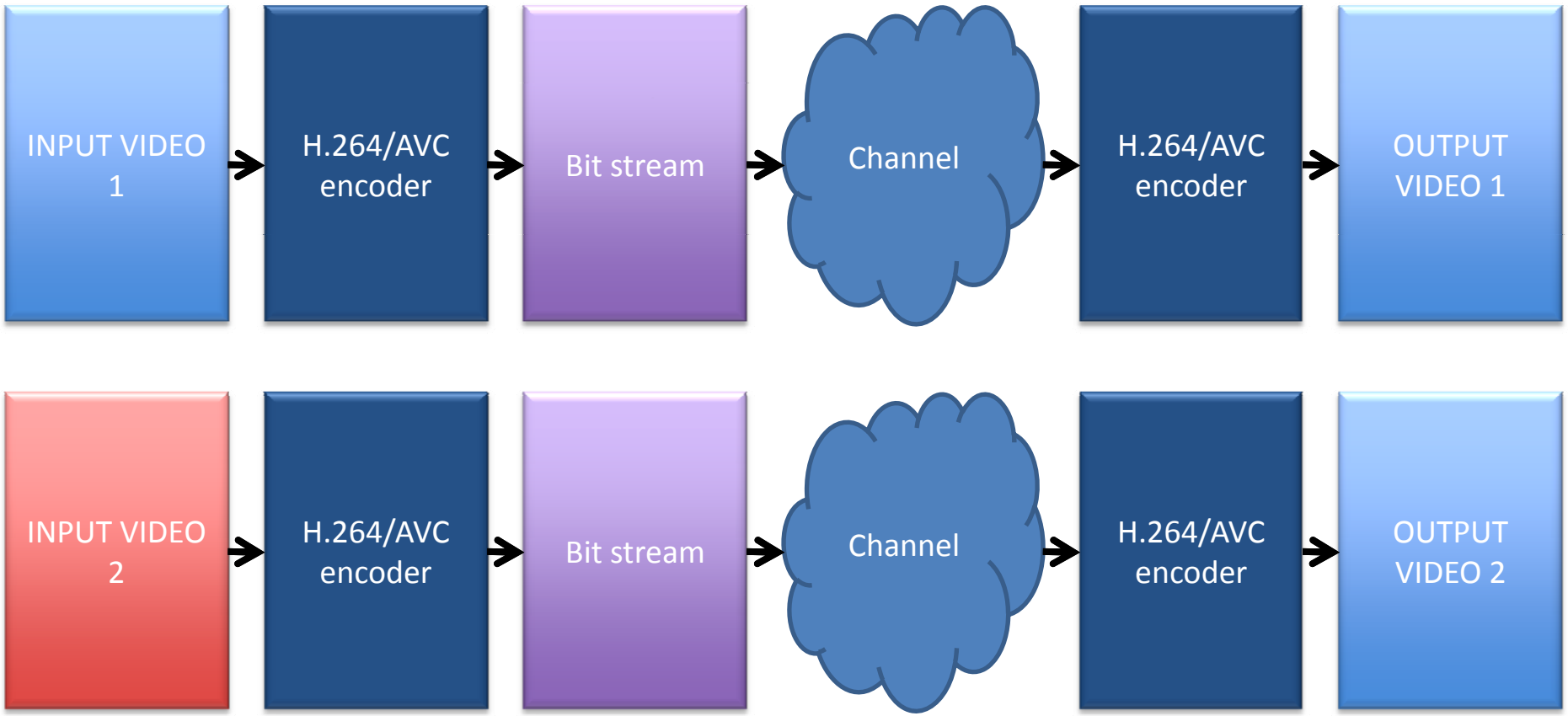
- **Transmission ready (DVB, Blu-Ray, …)**
- **Coding**
  - Compatibility, market introduction, …
  - Many existing standards
  - Rate proportional to no. of views?

Concepts applicable to:

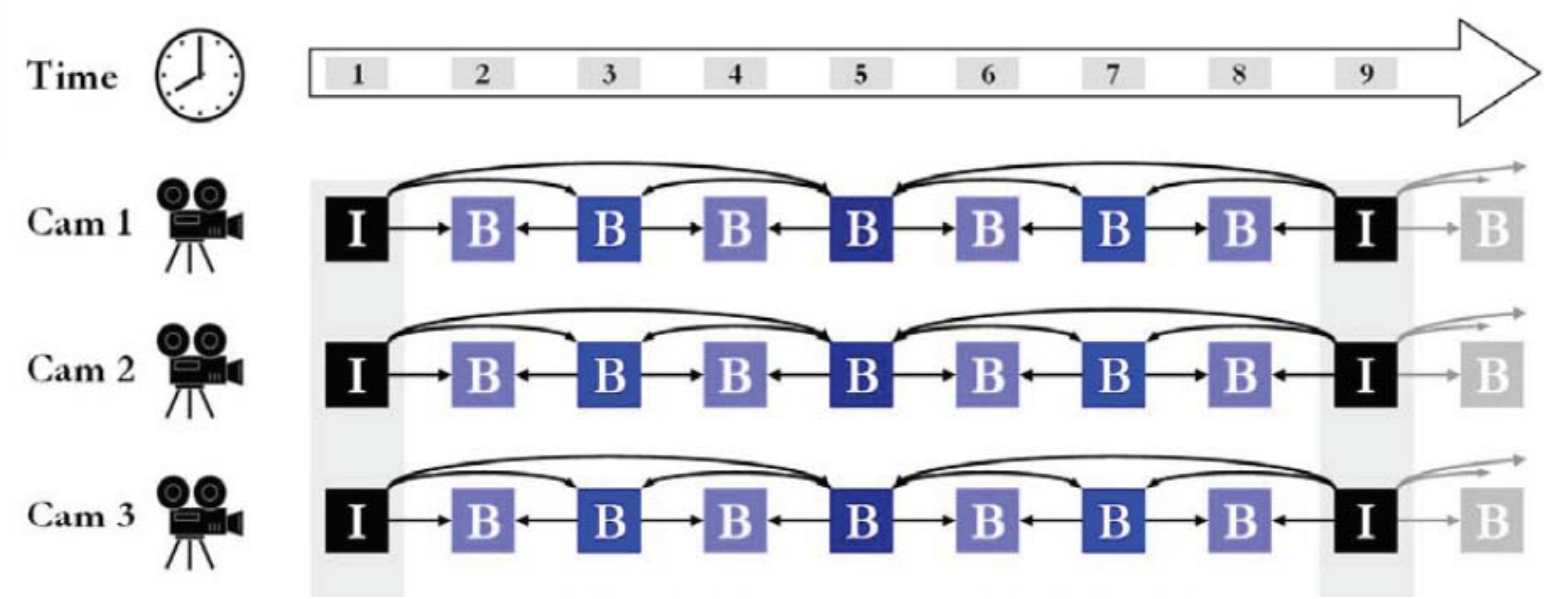
MPEG-1, **MPEG-2,** MPEG-4 Part 2, **AVC|H.264, HEVC**

- 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



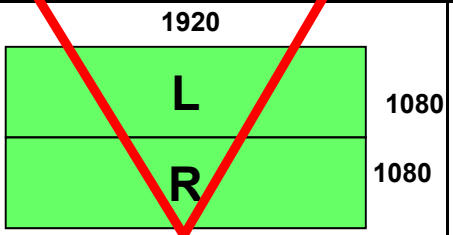
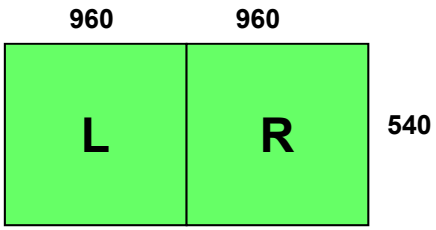
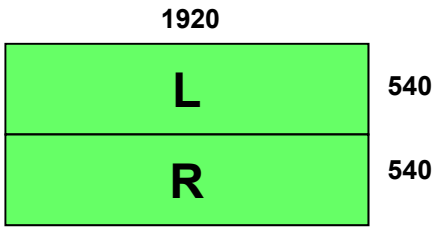
Frame Packing: Spatially Interleaved Views

Side-by-Side 960 x1080i

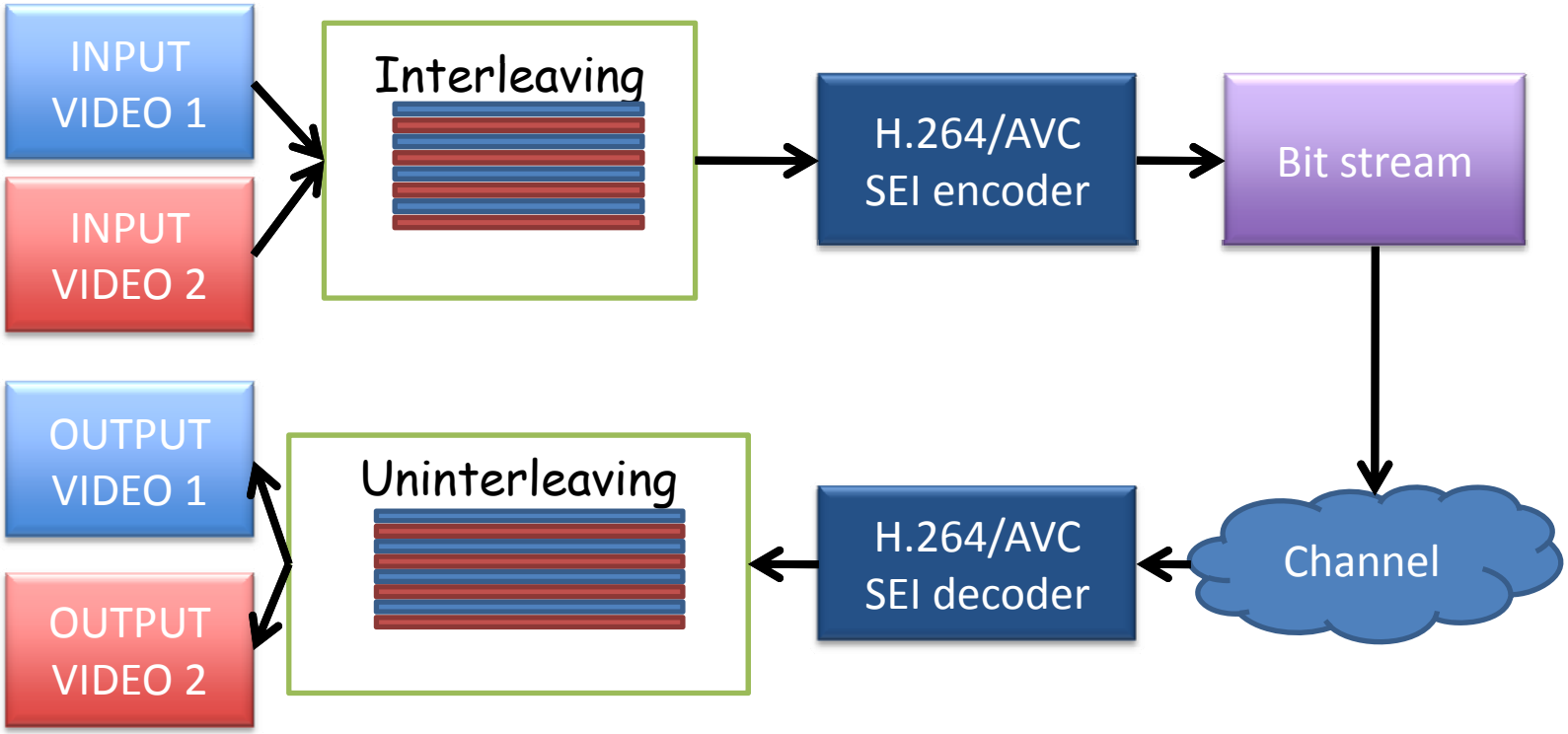
Over/Under 1920 x 540i

- **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

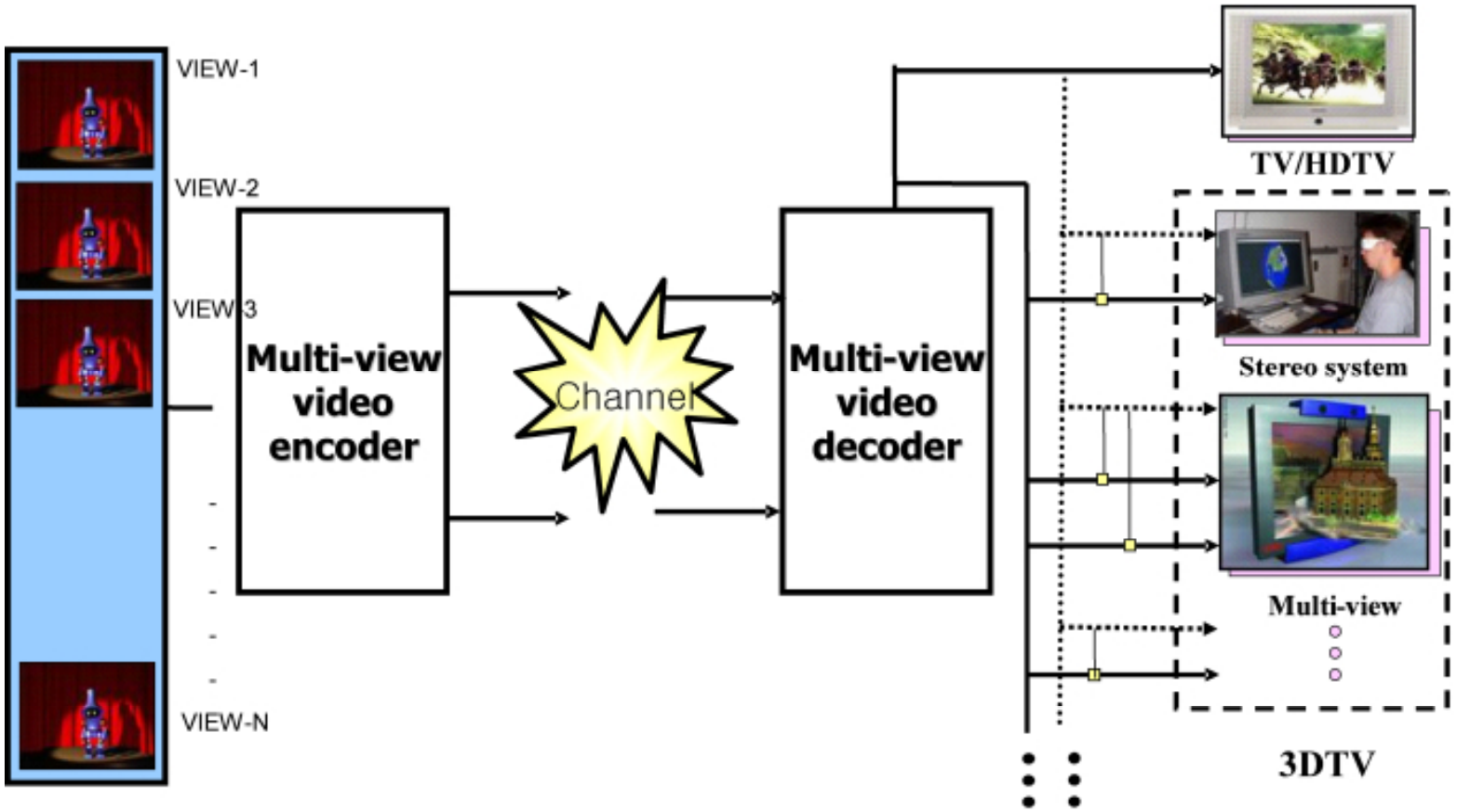
Side-by-Side: HDMI 1.4/1.3 Stereo-TV Formats

Frame Packing	Side-by-Side (Half)	Top-and-Bottom
		
l 1080p, 23.98 / 24 Hz l 720p, 59.94 / 60 Hz l 720p, 50 Hz	l 1080i, 59.94 / 60 Hz l 1080i, 50 Hz	l 1080p, 23.98 / 24 Hz l 720p, 59.94 / 60 Hz l 720p, 50 Hz
2 x HD	1 x HD	1 x HD

H.264 SEI Message

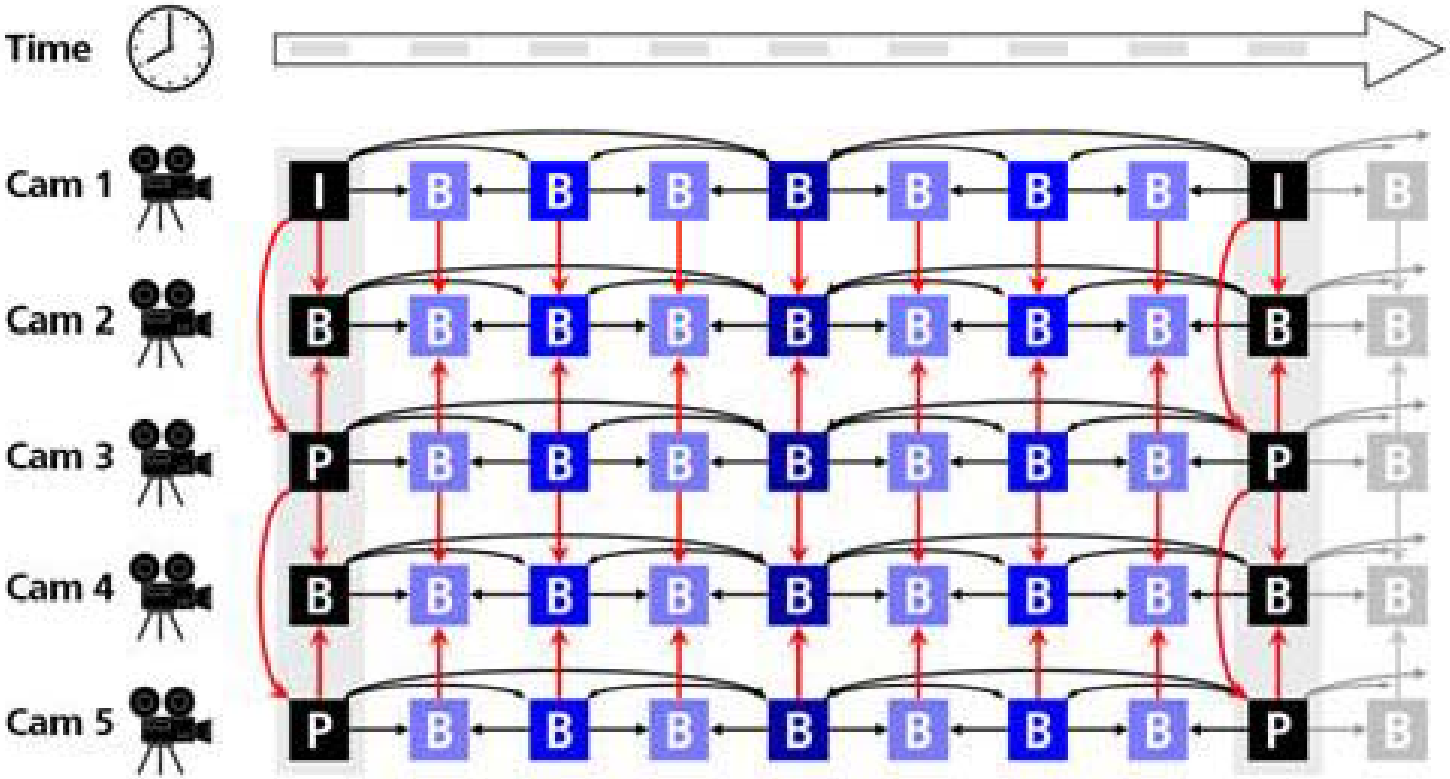


MVC: AVC Multi View Coding



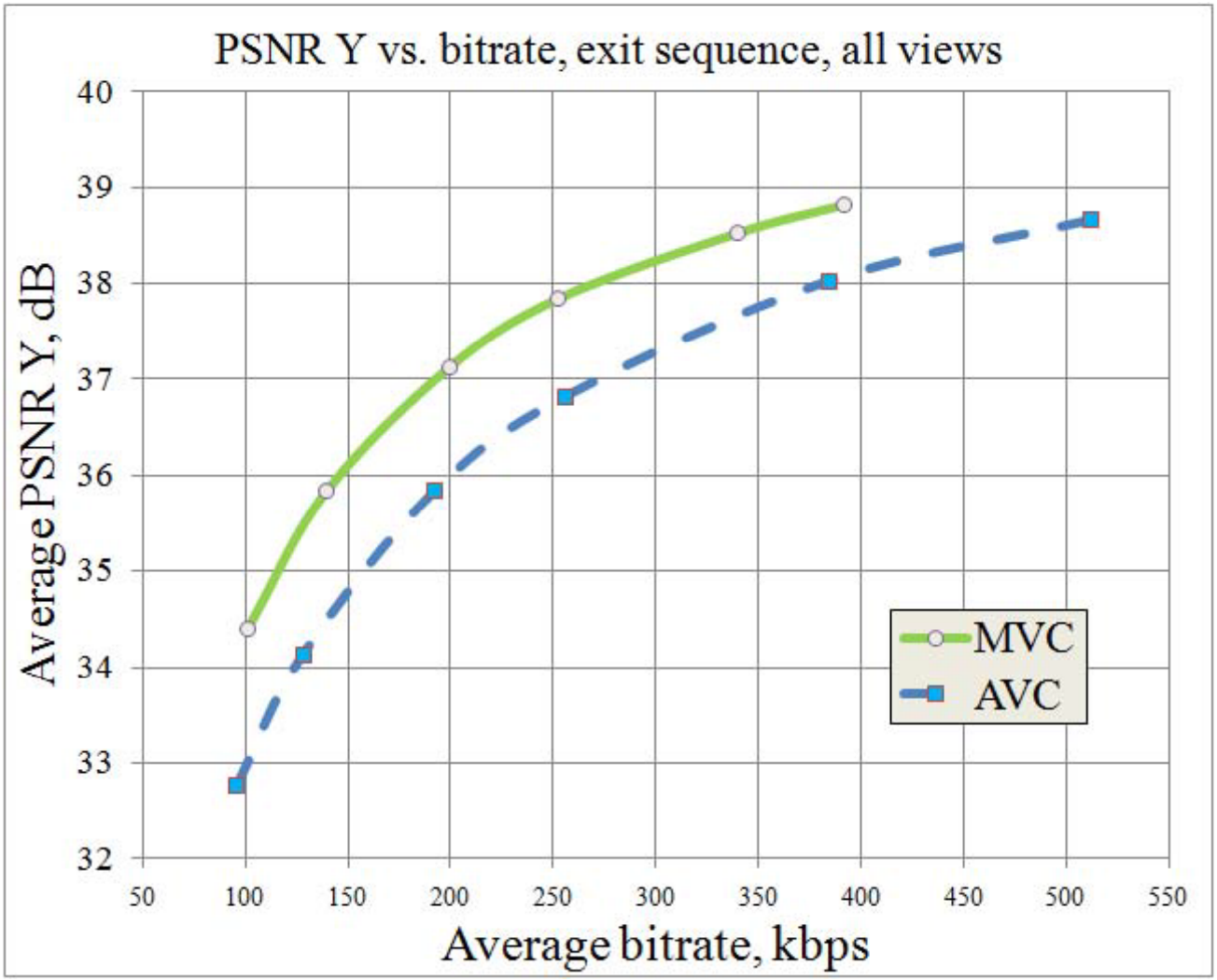
- **Defined as part of AVC/MPEG4/H.264**
  - ISO/IEC 14496-10:2008 Amendment 1

MVC: ISO/IEC 14496-10:2008 Amendment 1

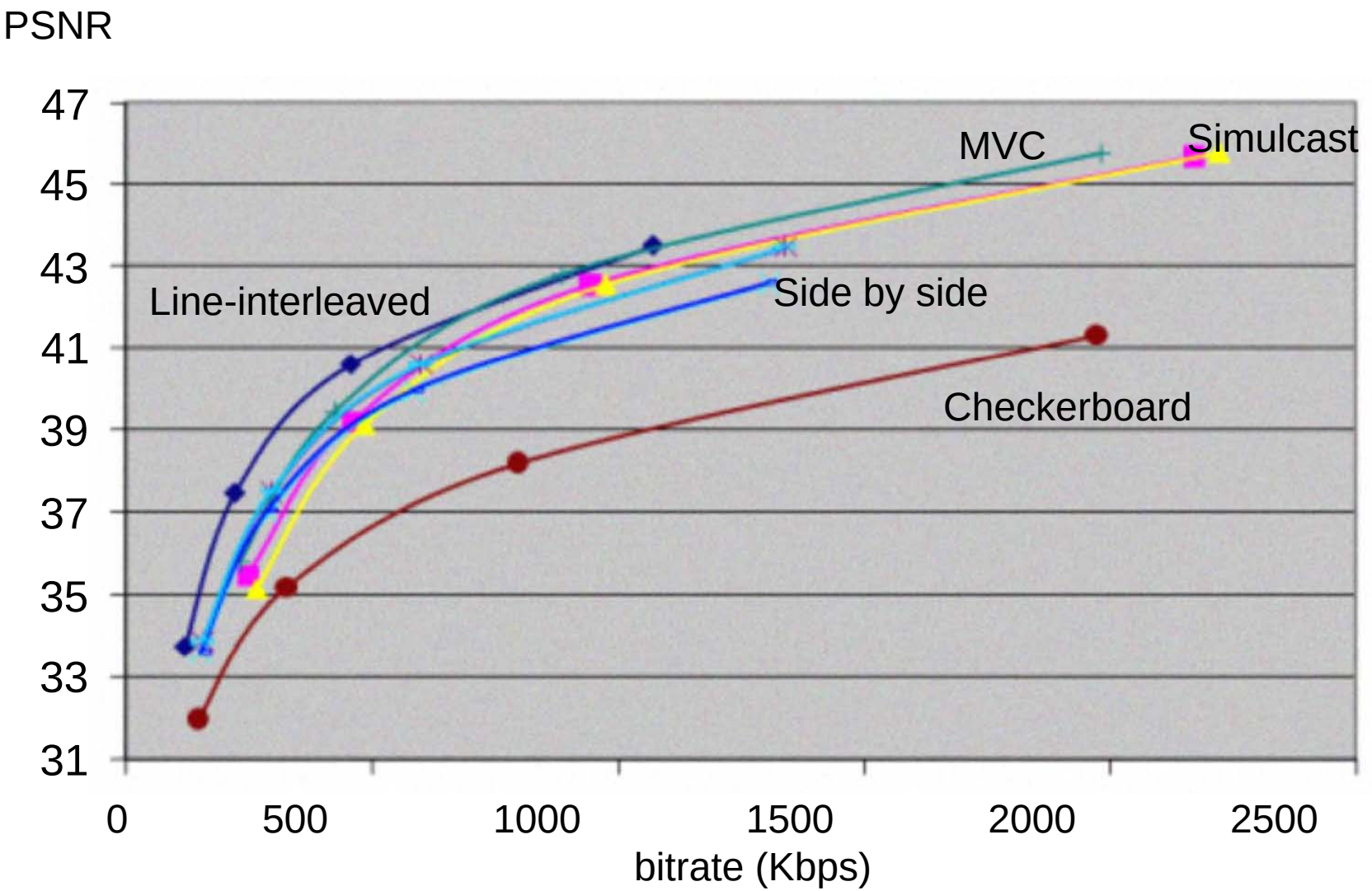


- Inter-frame and Inter-view (disparity) differences
- Based on standard AVC/H.264 compression techniques

Comparison AVC - MVC

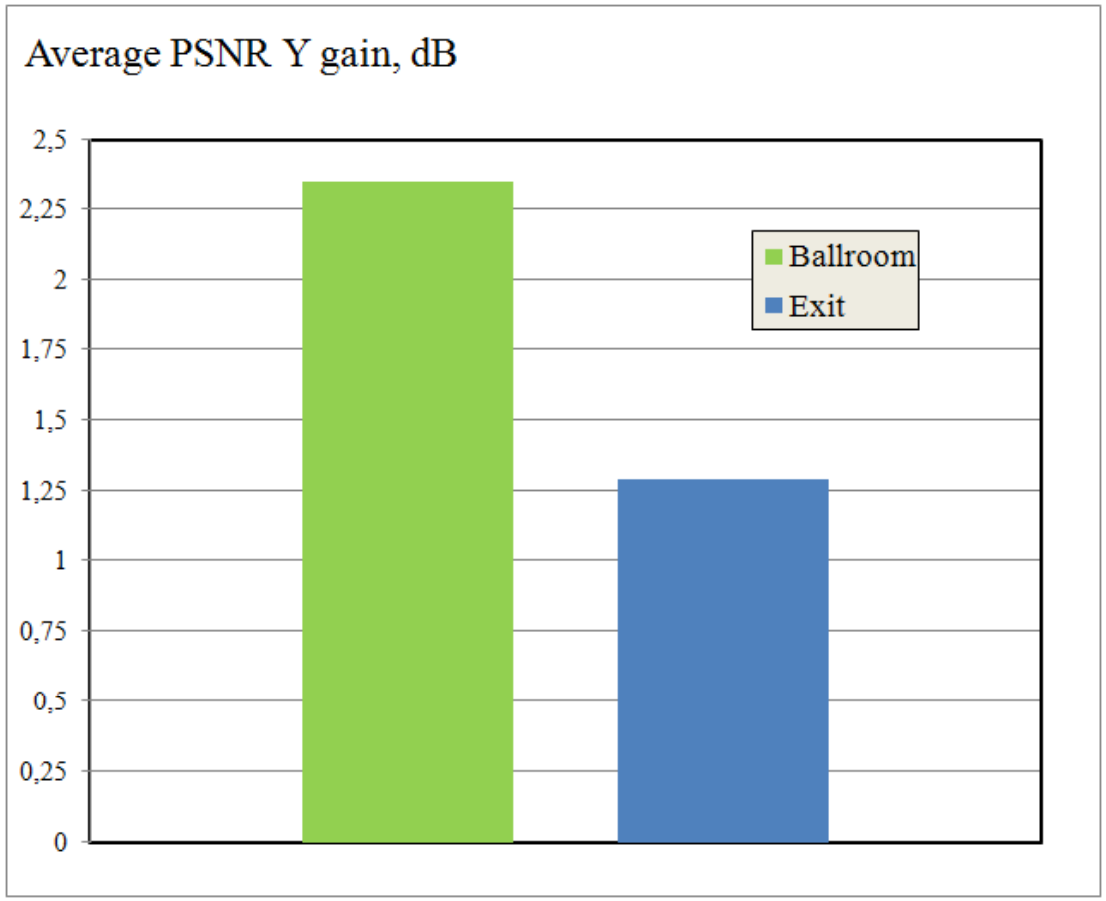


Efficiency of Frame Packing using AVC



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

Comparison AVC - MVC



50% rate increase compared to mono

Gain depends on number of views

- Up to 20%

2D Video plus Depth

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

Free Viewpoint TV: MPEG-C Part 3

– 1 or more videos plus depth map

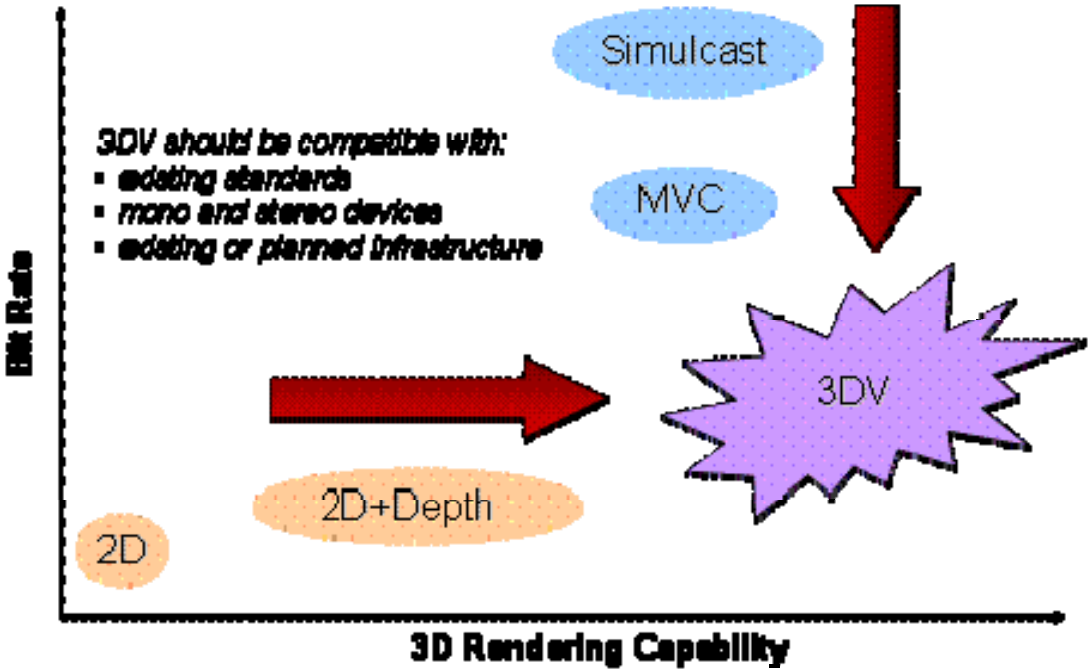
2D Video+Depth

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

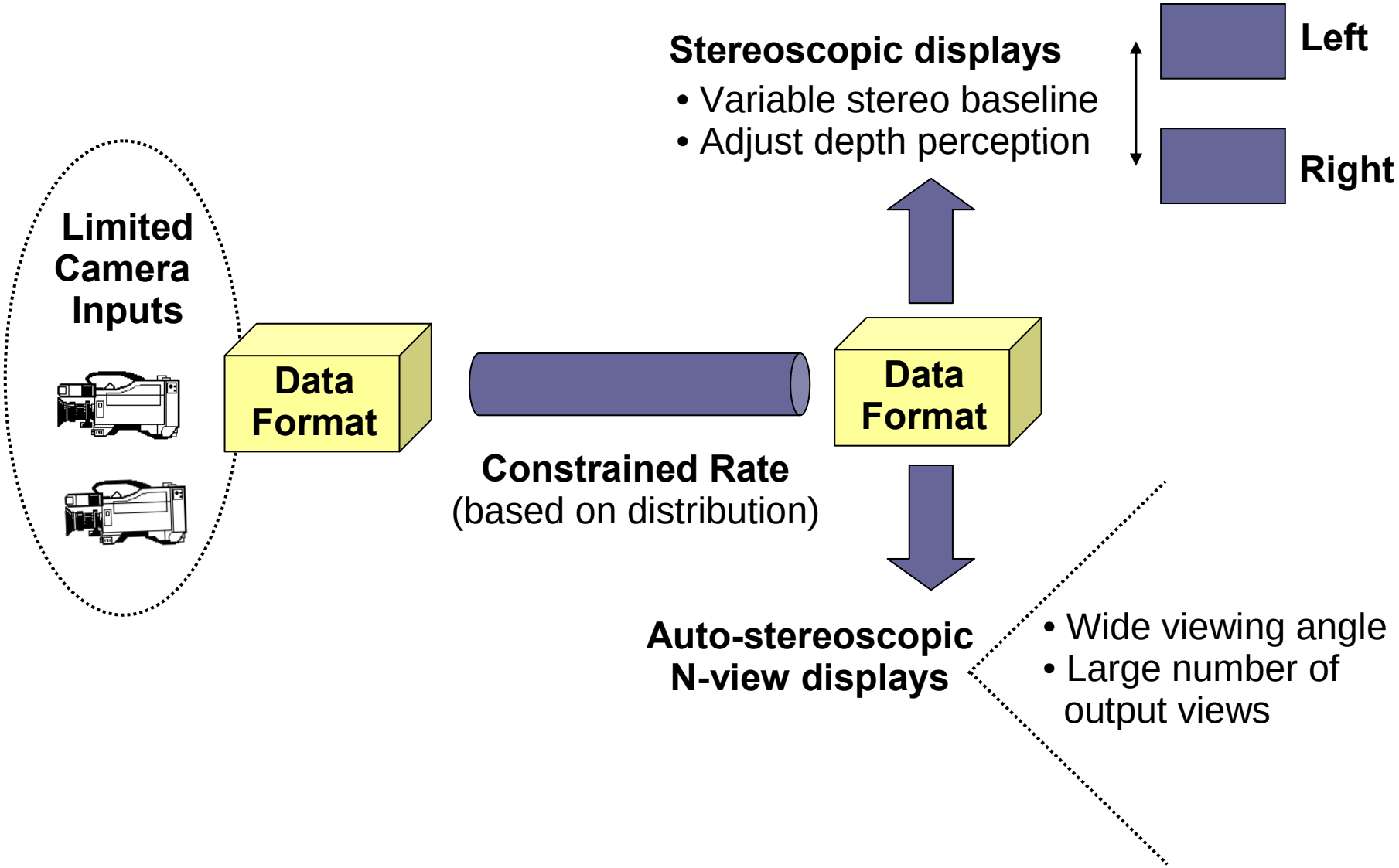
2DTV with Depth



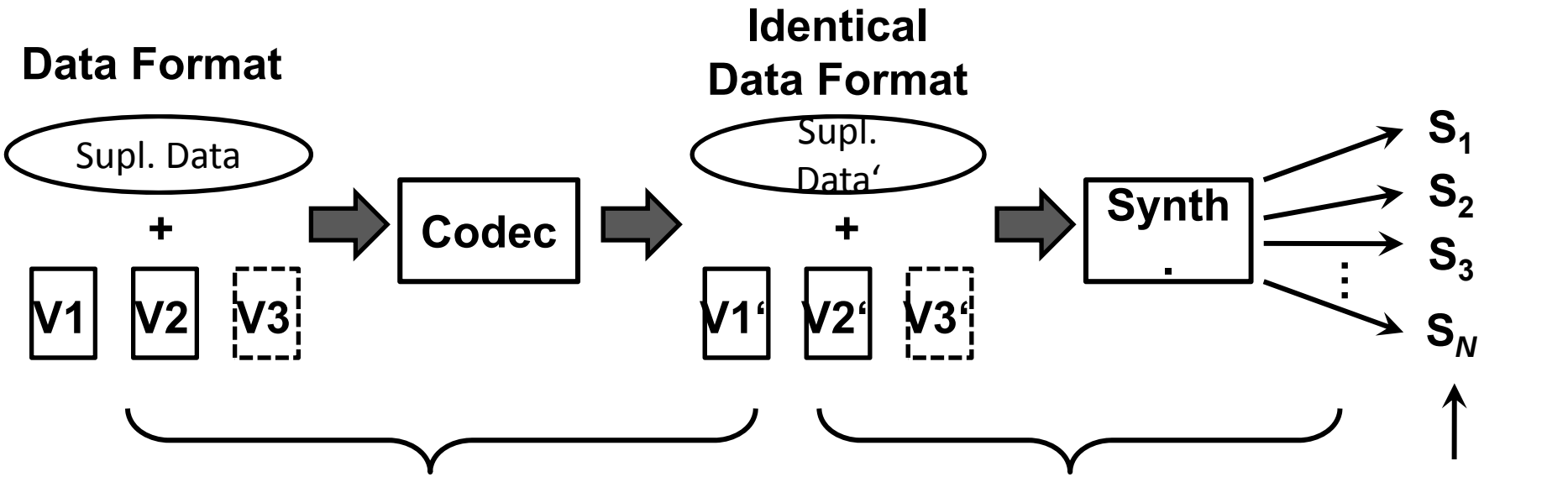
- **Depth Map sent as auxiliary video stream**
  - MPEG-C part 3 (ISO23002-3)
  - Depth Map computation difficult for natural scenes

Contents

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


Stereoscopic displays
• Variable stereo baseline
• Adjust depth perception
Left
Right
Limited Camera Inputs
Data Format
Constrained Rate
(based on distribution)
Data Format
Auto-stereoscopic N-view displays
• Wide viewing angle
• Large number of output views

3DV Evaluation



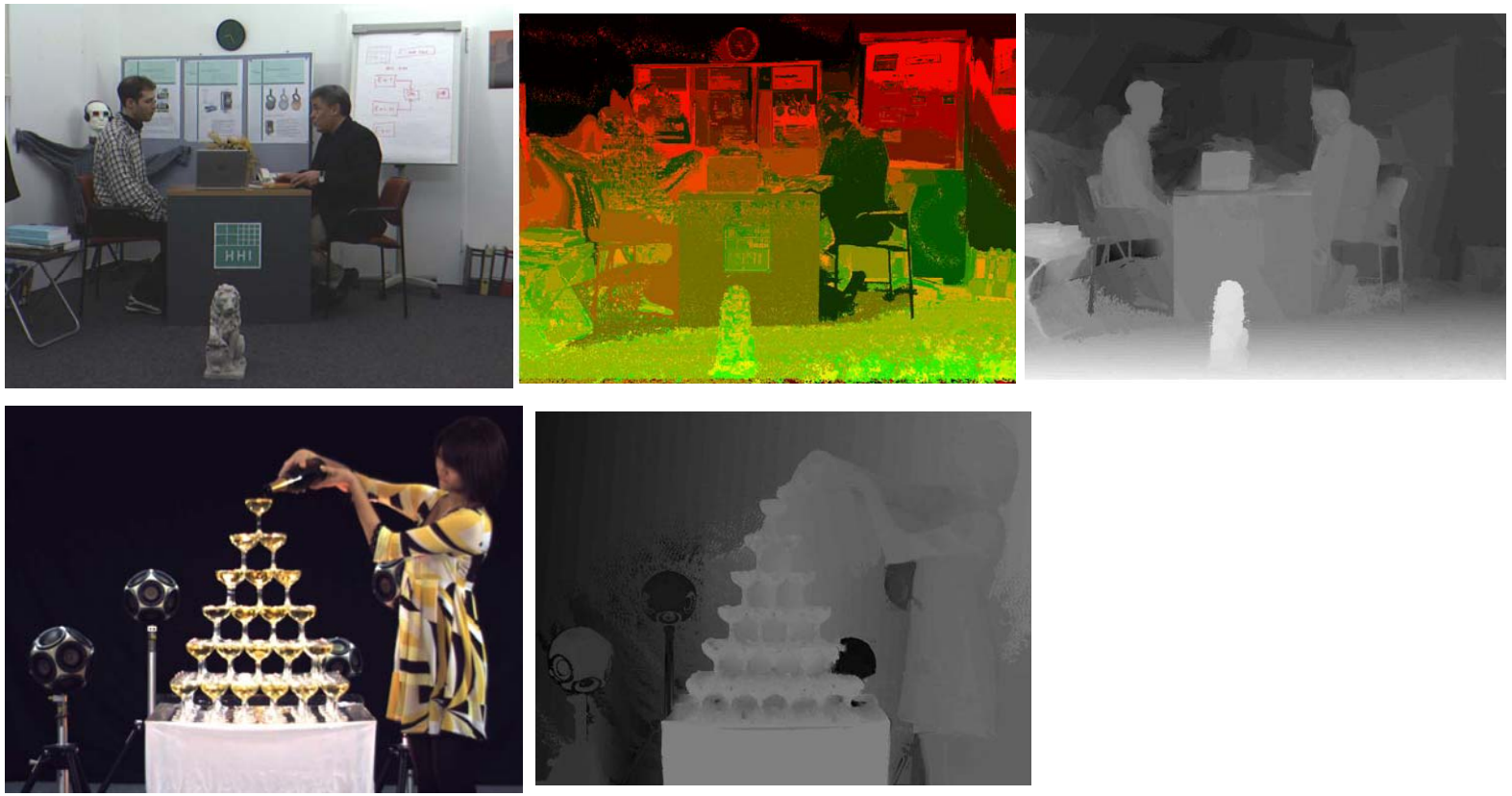
- One submission scenario, covering different data formats, compression, view synthesis
- Subjective 3D Evaluation of synthesized (random) stereo pairs / multi view
- Additionally, compression efficiency tested via automatic PSNR of video data and total overall bit rate (including all data format components)

Information for Proponents

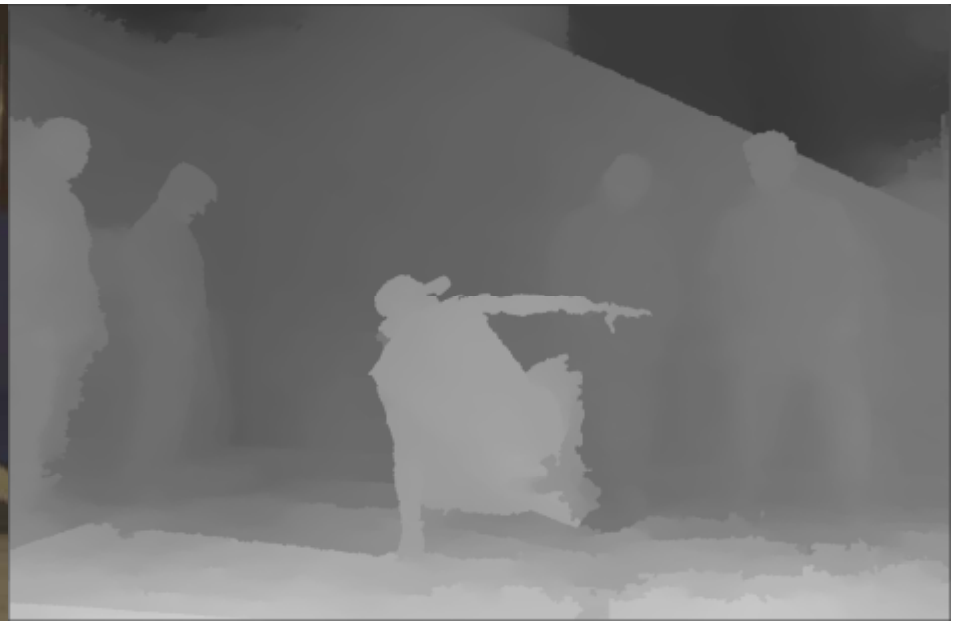
- **Compatibility**
  - AVC-compatibility: partial fulfillment of requirements, if base view is AVC-coded, second (and third) view coded with some new approach
  - HEVC-compatibility: Indicate HM version number
- **Data rate**
  - All data and parameters, required by a proponent's synthesis method to 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

Depth Estimation Reference Software (DERS)

- 3x3 block matching, depth recalculation using image segmentation and variable smoothing coefficient for graph cuts



View Synthesis Reference Software (VSRS)



$$Z = \frac{1}{\frac{Y}{255}\cdot\left(\frac{1}{Z_{near}}-\frac{1}{Z_{far}}\right)+\frac{1}{Z_{far}}}$$

View Synthesis Reference Software (VSRS)

- **Boundary aware hole filling**




- **Synthesis from left and right**



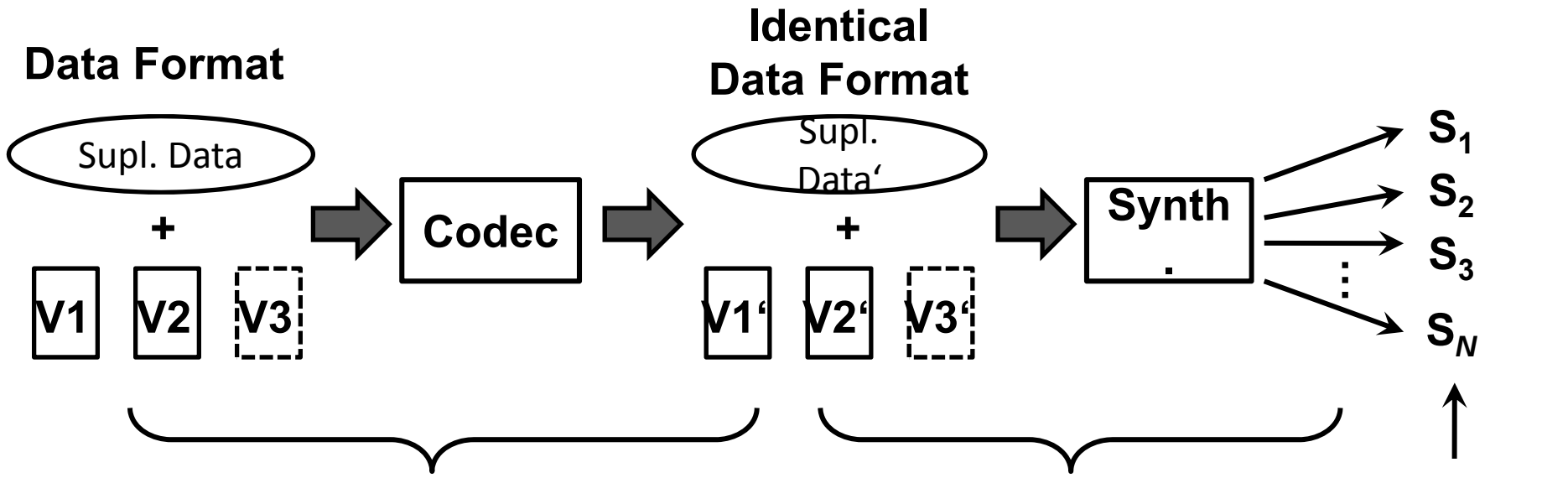
View Synthesis Reference Software (VSRS)

- Boundary noise removal

Exploration Experiments

- **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

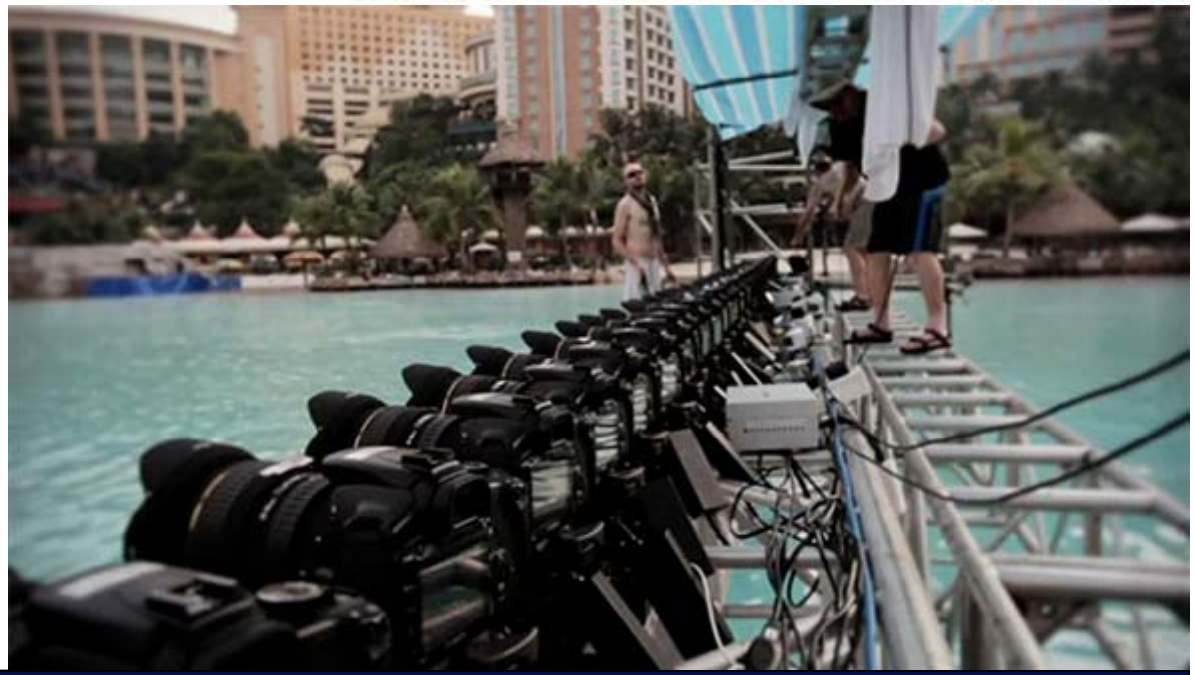
3DV Evaluation



- 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 A: 1920x1088p 25fps: "Poznan_Street", "Poznan_Hall2", "Undo_Dancer", "GT_Fly"
- 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

Seq. ID	Test Sequence	2-view input	3-view input
S01	Poznan_Hall2	7-6	7-6-5
S02	Poznan_Street	4-3	5-4-3
S03	Undo_Dancer	2-5	1-5-9
S04	GT_Fly	5-2	9-5-1
S05	Kendo	3-5	1-3-5
S06	Balloons	3-5	1-3-5
S07	Lovebird1	6-8	4-6-8
S08	Newspaper	4-6	2-4-6

Rates

Seq. ID	Test Sequence	2-view test scenario Bit rates (kbps)				3-view test scenario Bit rates (kbps)			
		R1	R2	R3	R4	R1	R2	R3	R4
S01	Poznan _Hall2	500	700	1000	1500	750	900	1300	2300
S02	Poznan _Street	500	700	1000	1250	750	1100	1800	4000
S03	Undo_ Dancer	1000	1300	1700	2200	1380	1750	2300	2900
S04	GT_Fly	1200	1700	2100	2900	2000	2380	2900	4000
S05	Kendo	400	500	800	1300	800	1000	1300	1900
S06	Balloons	320	430	600	940	500	600	800	1250
S07	Lovebird1	375	500	750	1250	500	800	1250	2000
S08	News-paper	400	525	800	1300	500	700	1000	1350

View Synthesis

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

Tool Categories

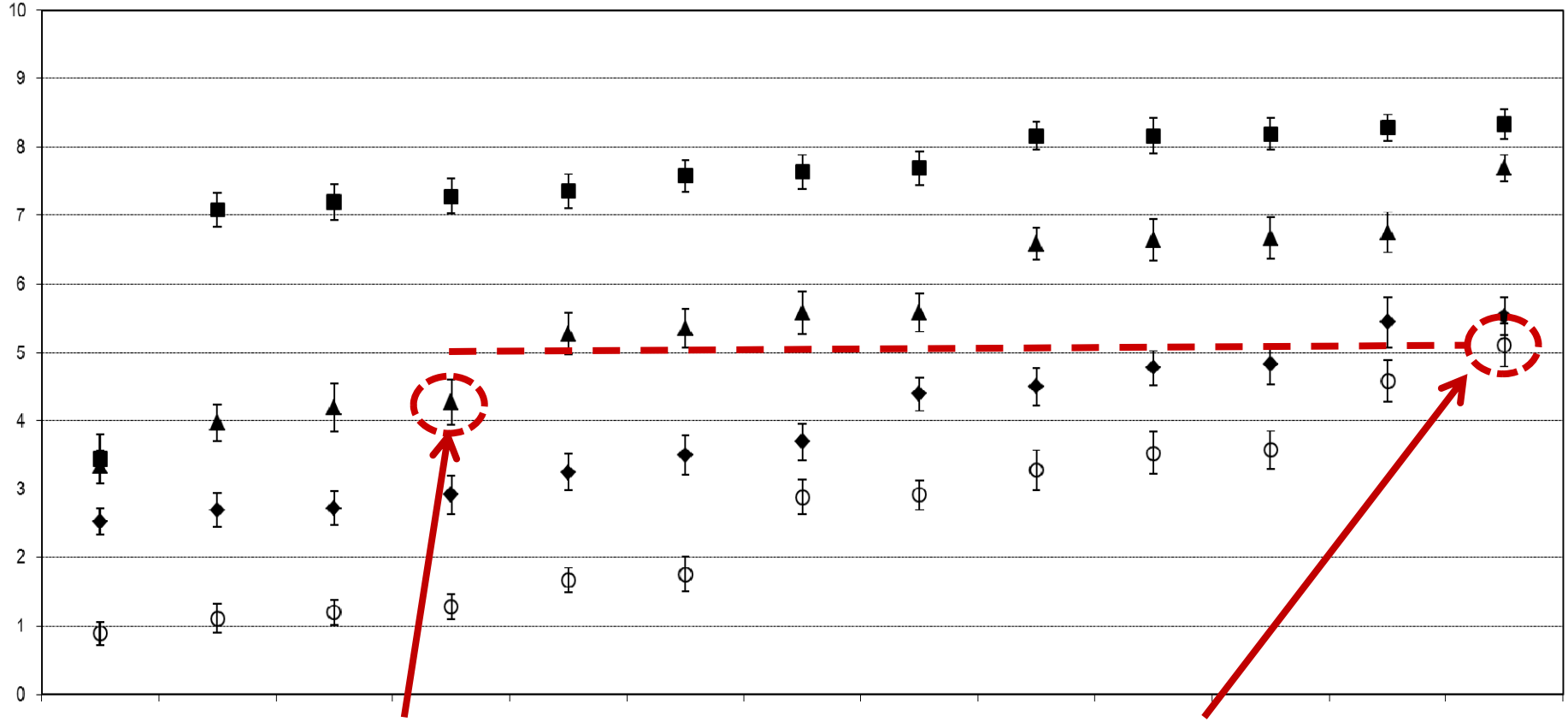
- **Texture coding**
  - Independent of depth
    - E.g., inter-view prediction of color view, inter-view prediction of motion parameters and residual data
  - Using depth data
    - E.g., view synthesis prediction, motion prediction, motion-compensated prediction, QP selection, inloop filter
- **Depth coding**
  - Independent of texture
    - E.g., depth modeling modes, weighted prediction
  - Using texture data
    - E.g., motion parameter inheritance, intra prediction

Subjective Tests (11/2011)

- **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

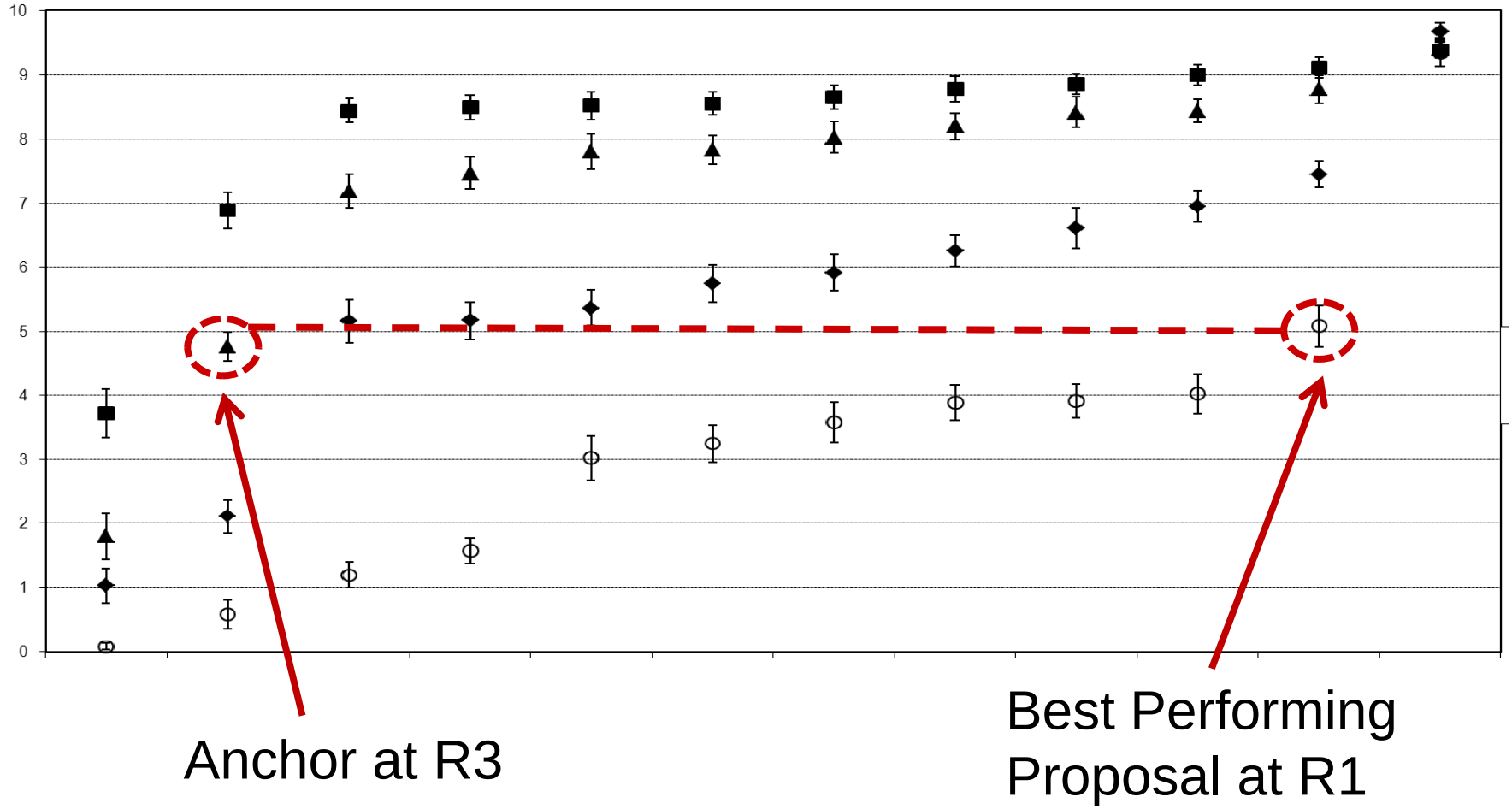


Anchor at R3

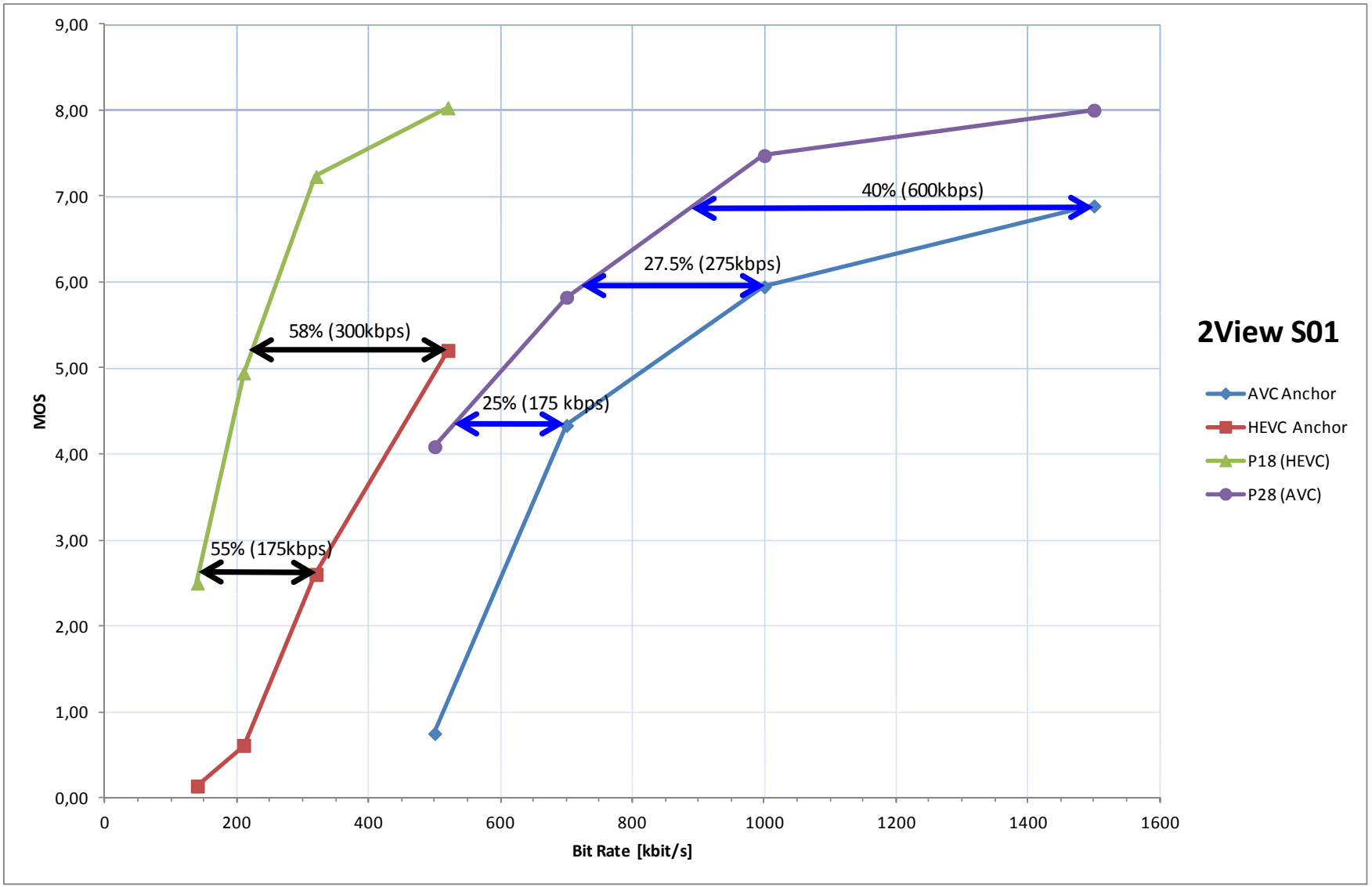
Best Performing
Proposal at R1

Sample Subjective Test Result

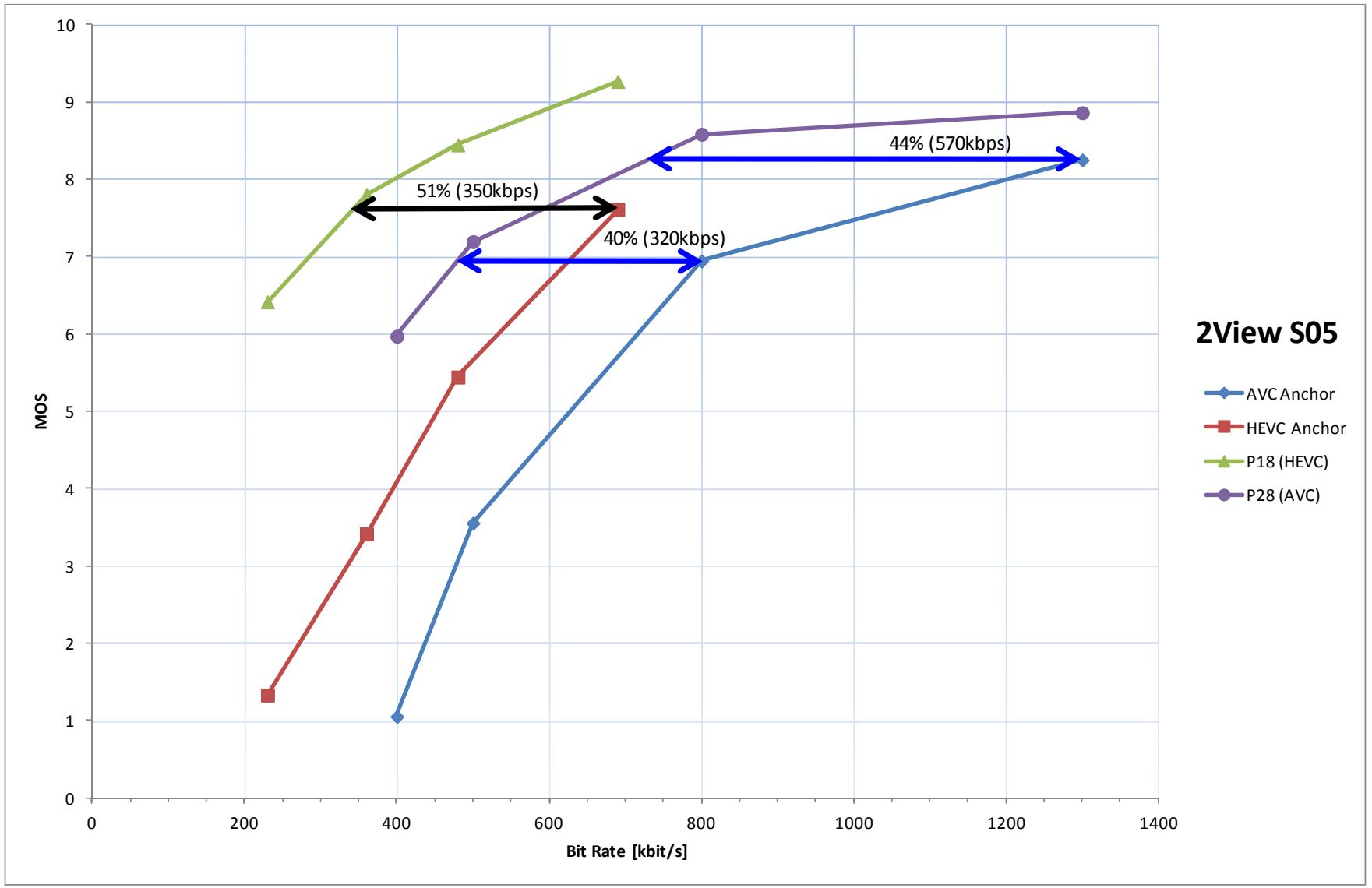
Class A Sequence, 2-view test case
HEVC-Compatible



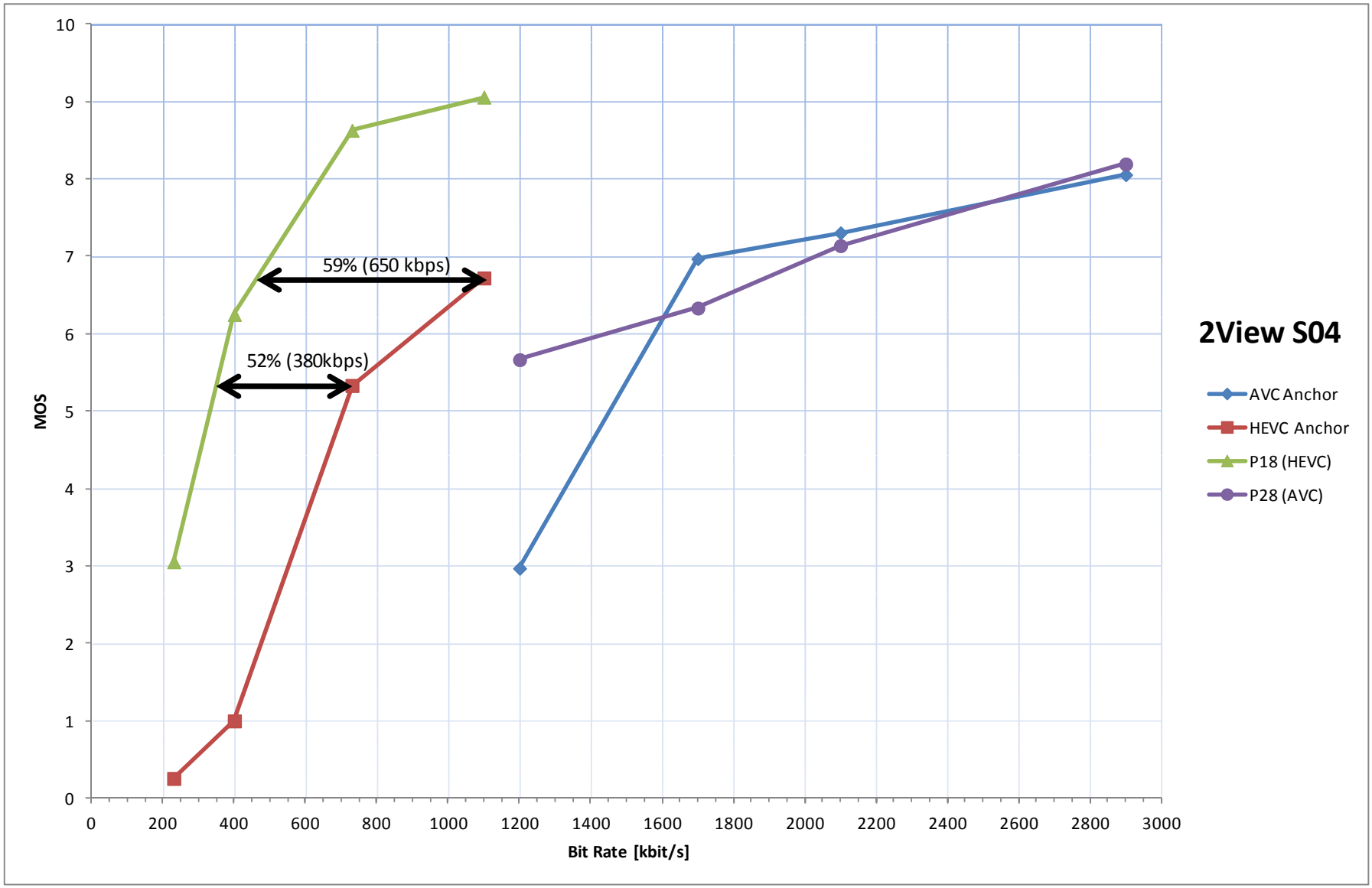
2 View Scenario



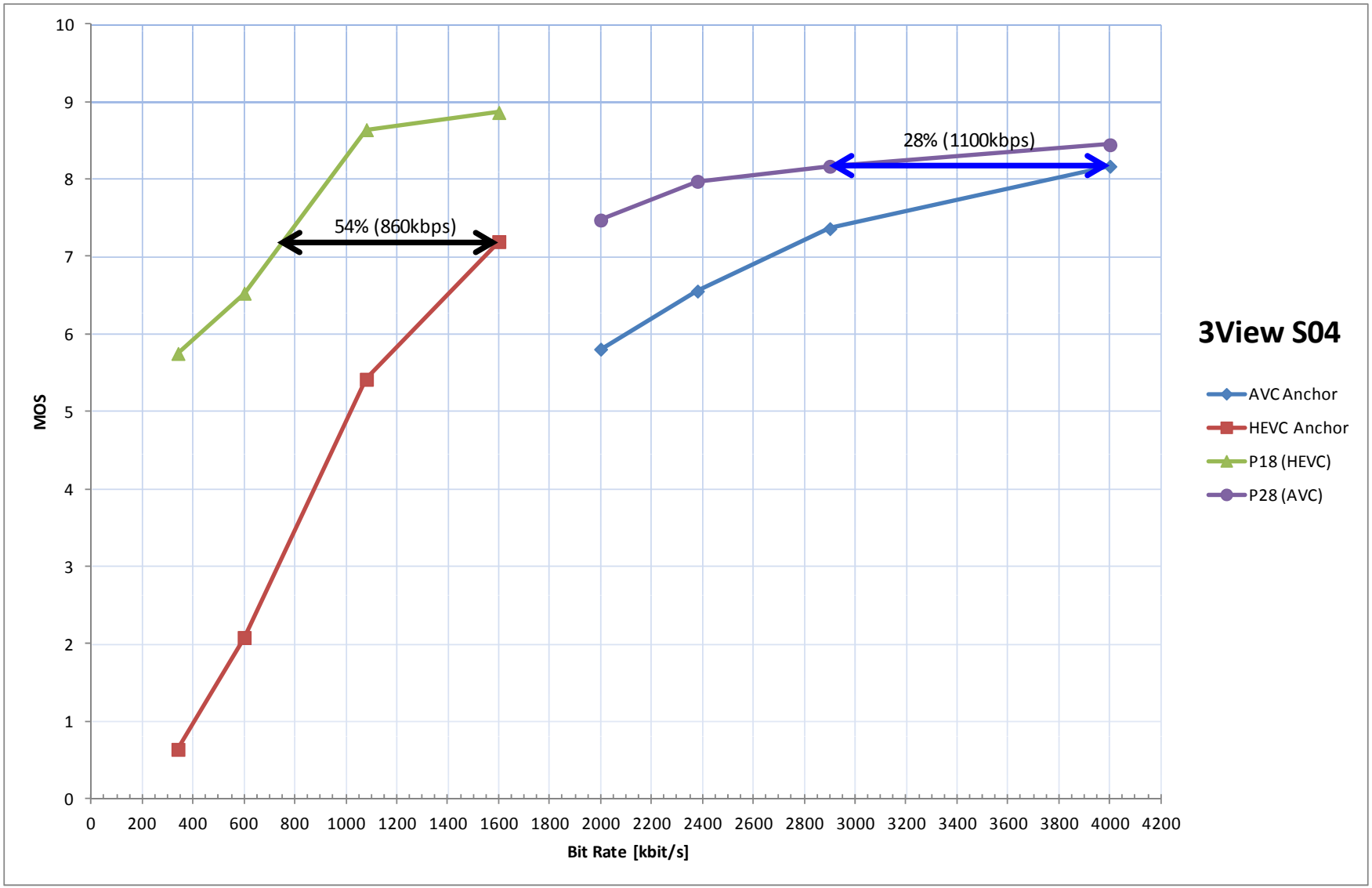
2 View Scenario



2 View Scenario



3 View Scenario



- 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

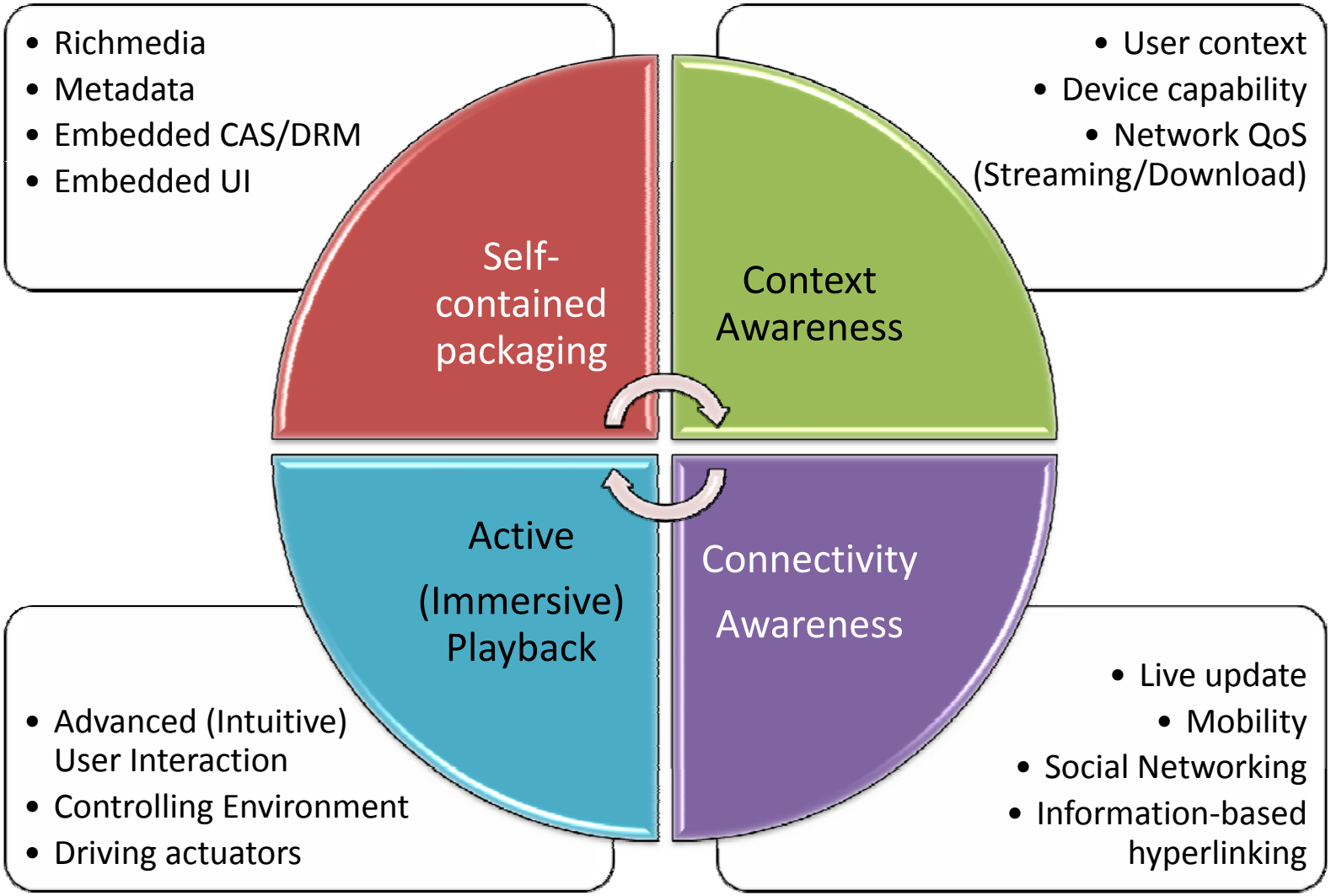
- One approach based on mesh-based decoder-side view synthesis

- **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)

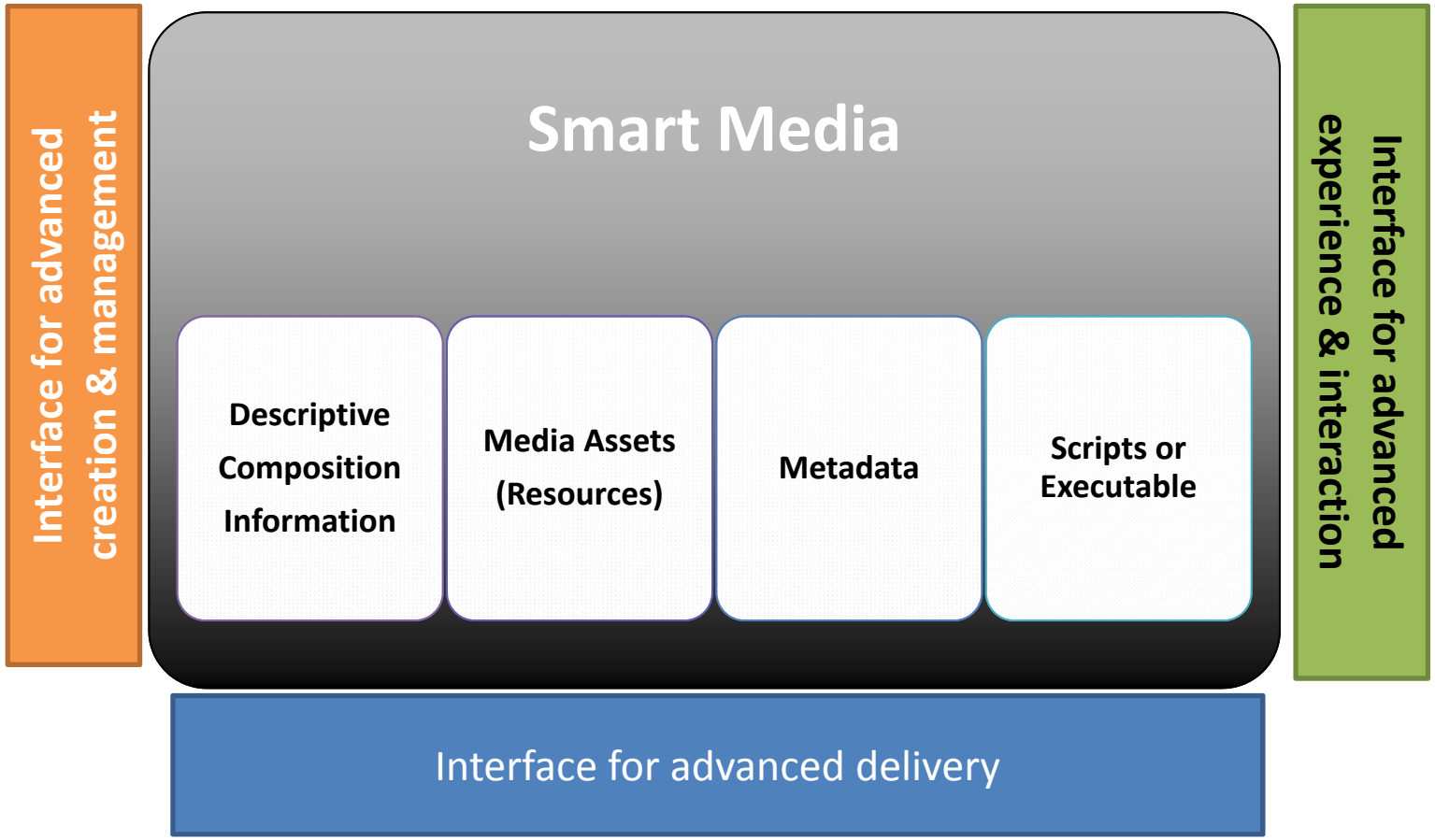
Contents

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

MPEG-H: Standards for SmartMedia Technologies



MPEG-H: Standards for SmartMedia Technologies



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

JCTVC-G399:HEVC (HM 4.0) vs. AVC (Strong JM) based on PSNR

	Random Access	Low Delay	All Intra
Class A	43%		29%
Class B	44%	48%	26%
Class C	34%	41%	23%
Class D	32%	38%	18%
Class E		51%	29%
Average	**39%**	**44%**	**25%**

Potential other Standards for 3DV

- MPEG-2 plus HEVC
- AVC plus HEVC

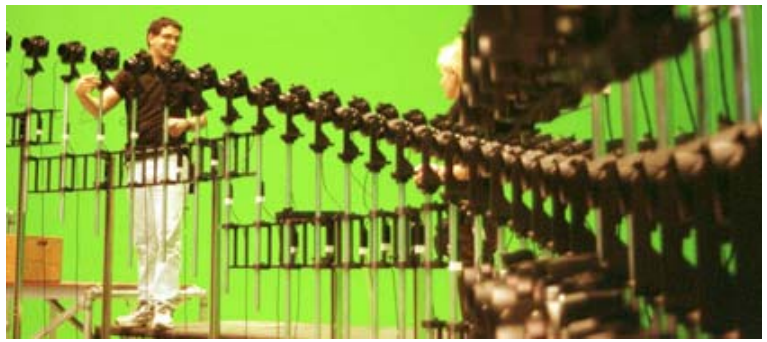
- 3DV based on HEVC
- 3DV based on AVC

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?
- …



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