MPEG Developments in Multi-view Video Coding and 3D Video

Jens-Rainer Ohm

RWTH Aachen University
Lehrstuhl und Institut für Nachrichtentechnik
ohm@ient.rwth-aachen.de
http://www.ient.rwth-aachen.de

Outline

1. Introduction – Purpose and Applications
2. Stereo and Multi-view Video Coding standardization in MPEG and JVT
3. 2D Video plus depth (MPEG-C part 3)
4. 3D Video / Free-viewpoint Video
5. Conclusions

Stereo and Multi-view Video Coding

- "**Classic" Stereo** requires only two views which are taken "as is" – i.e. the capture must already take display properties into account
- **Compression of stereo video** is straightforward
  - Simulcast
  - Combination of two views into one
  - Exploitation of inter-view redundancy

- This does not support
  - **N-view displays** (autostereoscopic, holographic)
  - Additional functionality: **Baseline adaptation**
- For these purposes, either **coding of multiple views** (if available) or **depth-based synthesis** is needed

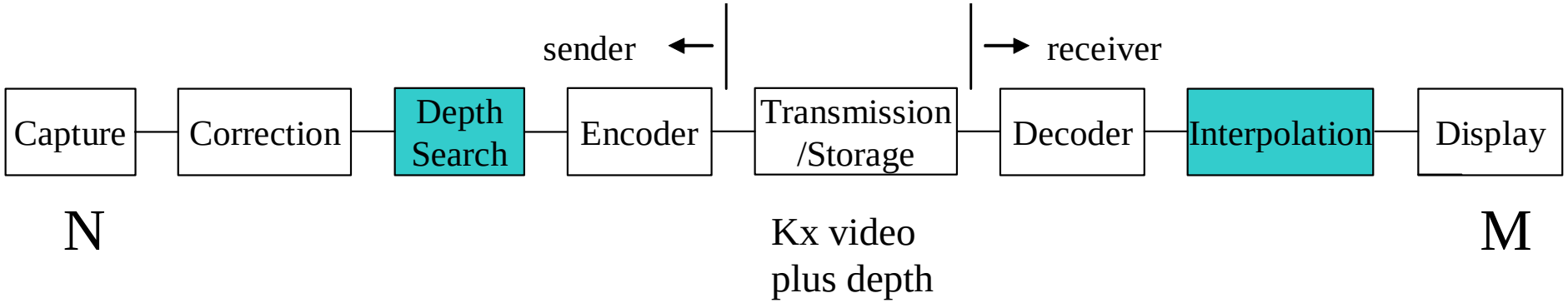
Multi-view Video Coding (MVC)

- Multi-view and 3D video representations require **multiple synchronized video signals** that show the same scenery from different viewpoints
- **Huge amount of data** with need to be compressed efficiently
- Multiview typically has a larger amount of **inter-view statistical dependencies** than stereo

We would like to thank the Image Based Realities Group of Microsoft Research for providing the Breakdancers and Ballroom data sets.

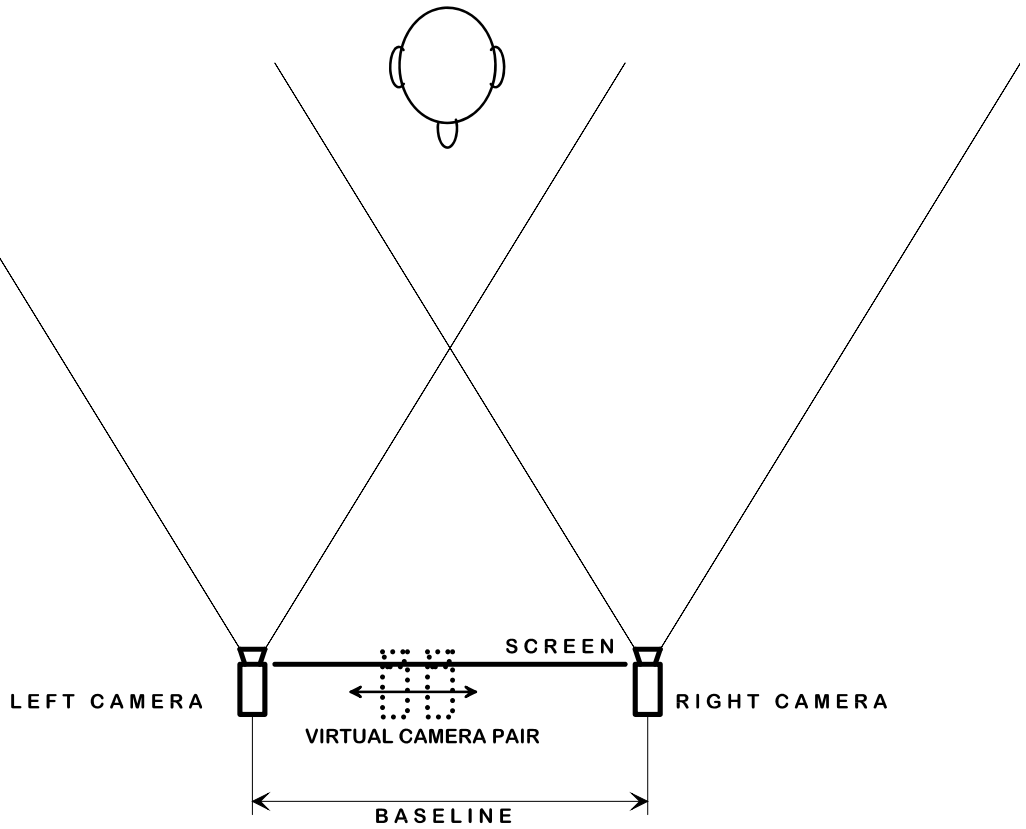
Video-plus-Depth / 3D Video

- Support for **N-view displays** (various types) or **stereo with baseline adaptation** where only low amount of views (1-3) and associated depth map(s) is encoded
  - Generate **synthesized views** using **video and depth**
  - At minimum: One video, one depth map
- **Technologies required:**
  - Depth estimation (non-normative)
  - Depth encoding (normative)
  - View synthesis (non-normative or with minimum normative quality requirements)



Video-plus-Depth / 3D Video

- Examples:
  Stereo with baseline adaptation, or view shift (e.g. after head tracking)

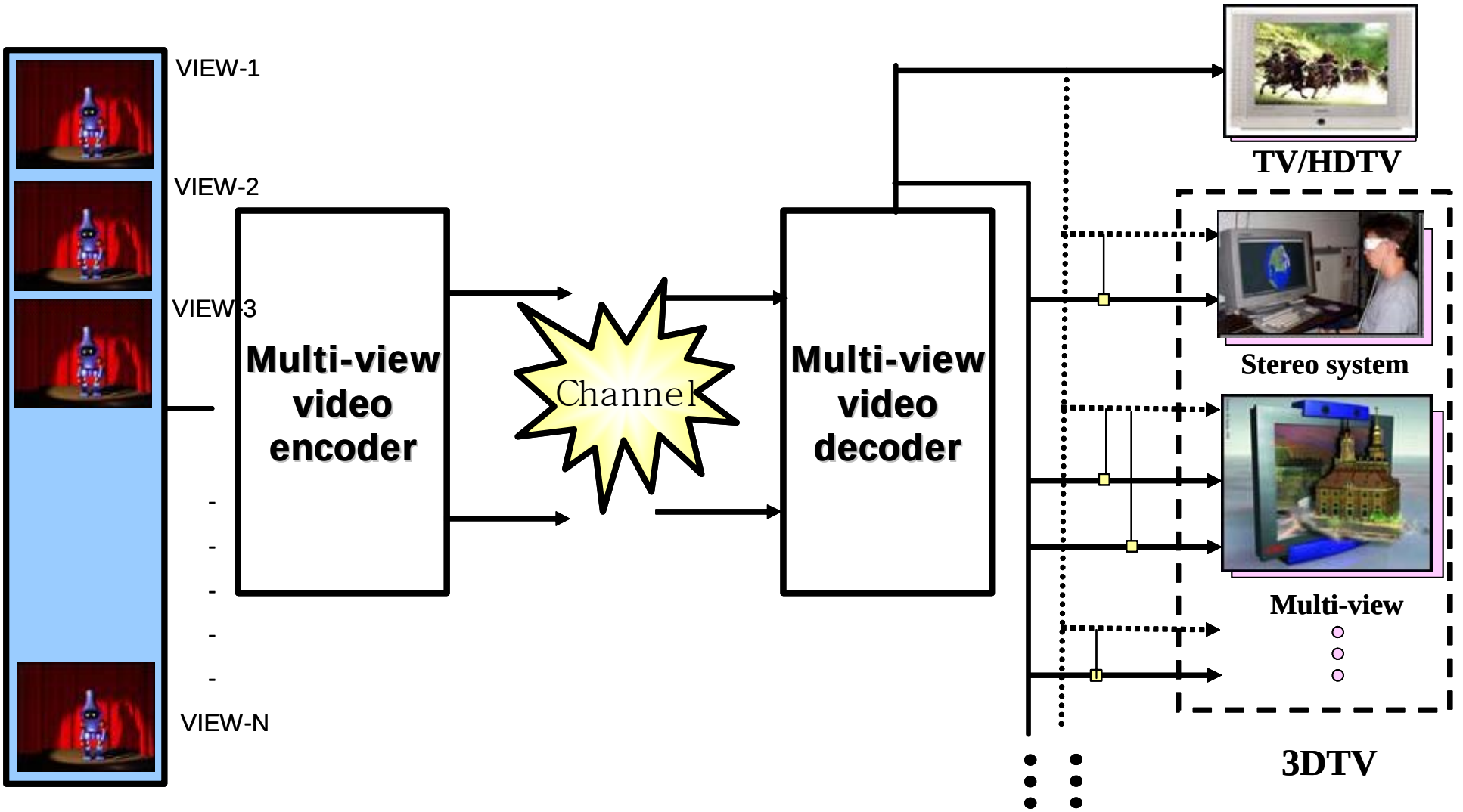


- Maximum angle between leftmost and rightmost position expected to be around 20 degrees – also for the upcoming generations of N-view autosteroscopic displays

Stereo Video

- **L/R simulcast** possible with any MPEG standard
- **MPEG-2 Multi-view profile** is essentially stereo with temporal L/R interleaving
- **Stereoscopic MAF** ISO/IEC 23000-11 based on MPEG-4 part 2 video (L/R packing, for handhelds)
- **MPEG-4 part 10 AVC** Stereo SEI message and Frame Packing Arrangement SEI message (the latter in 14496-10/5e Amd.1, to be finalized by July 2009) allow **various methods of L/R packing**
  - Temporal, spatial row/column, spatial side-by-side/up-and-bottom, checkerboard (quincunx)
- **MPEG-4 AVC Stereo High Profile** (new in Study 14496-10/5e Amd.1, to be finalized by July 2009)
  - Subset of MVC, restricted to 2 views, allows progressive and interlaced stereo

Multi-view Video Coding (MVC)



MVC Standard – Status and Overview

Standard was approved in July 2008

- Specified as an amendment of H.264/MPEG-4 AVC
- Integrated into 5th Edition of ISO/IEC 14496-10 (Annex H)

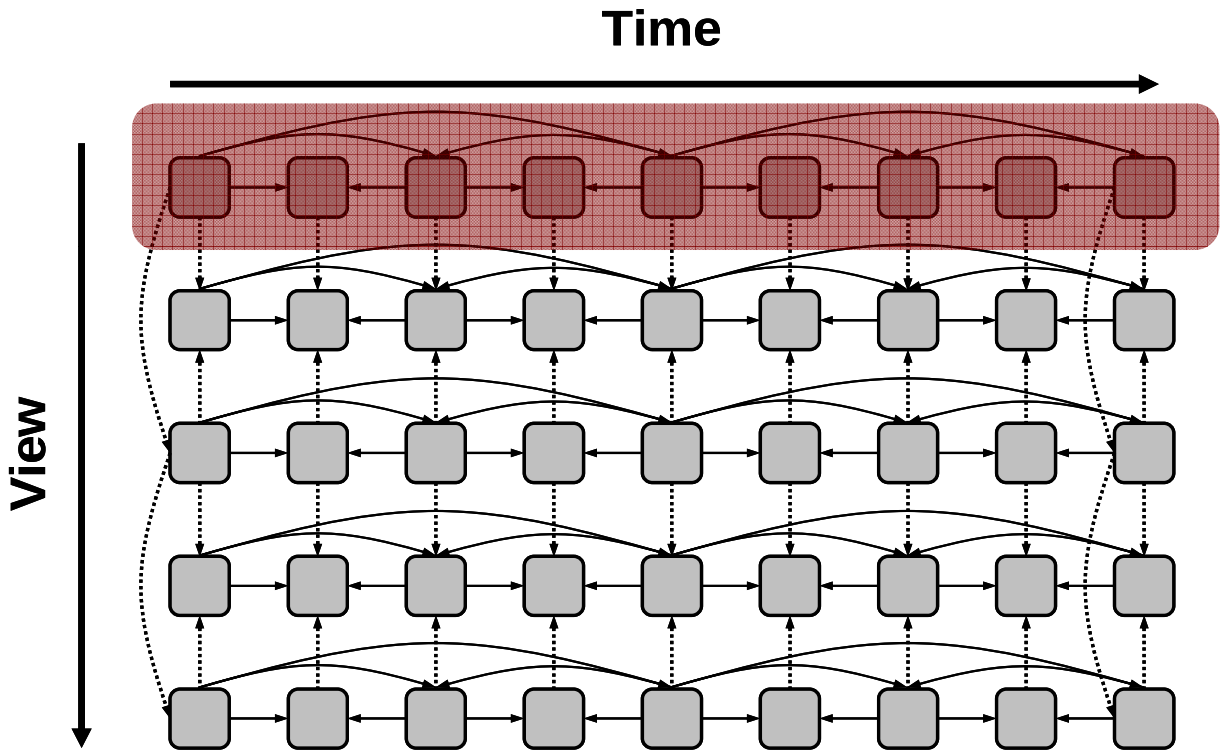
Key Elements of MVC Design

- **Syntax**
  - No changes to lower-level AVC syntax (slice and lower), so compatible and easily integrated with existing hardware
  - Small backward compatible changes to high-level syntax, e.g., to specify view dependency, random access points
  - Base layer required and easily extracted from video bitstream (identified by NAL unit type syntax)
- **Inter-view prediction**
  - Enabled through flexible reference picture management
  - Allow decoded pictures from other views to be inserted and removed from reference picture buffer
  - Core decoding modules not aware of whether reference picture is a time reference or view reference

MVC Standard – Inter-view Prediction

Prediction across views to exploit inter-camera redundancy

- Dependencies flexible for multiview, much simpler for stereo
- Limitations: (a) inter-view prediction only from same time instance
  (b) cannot exceed maximum number of stored reference pictures

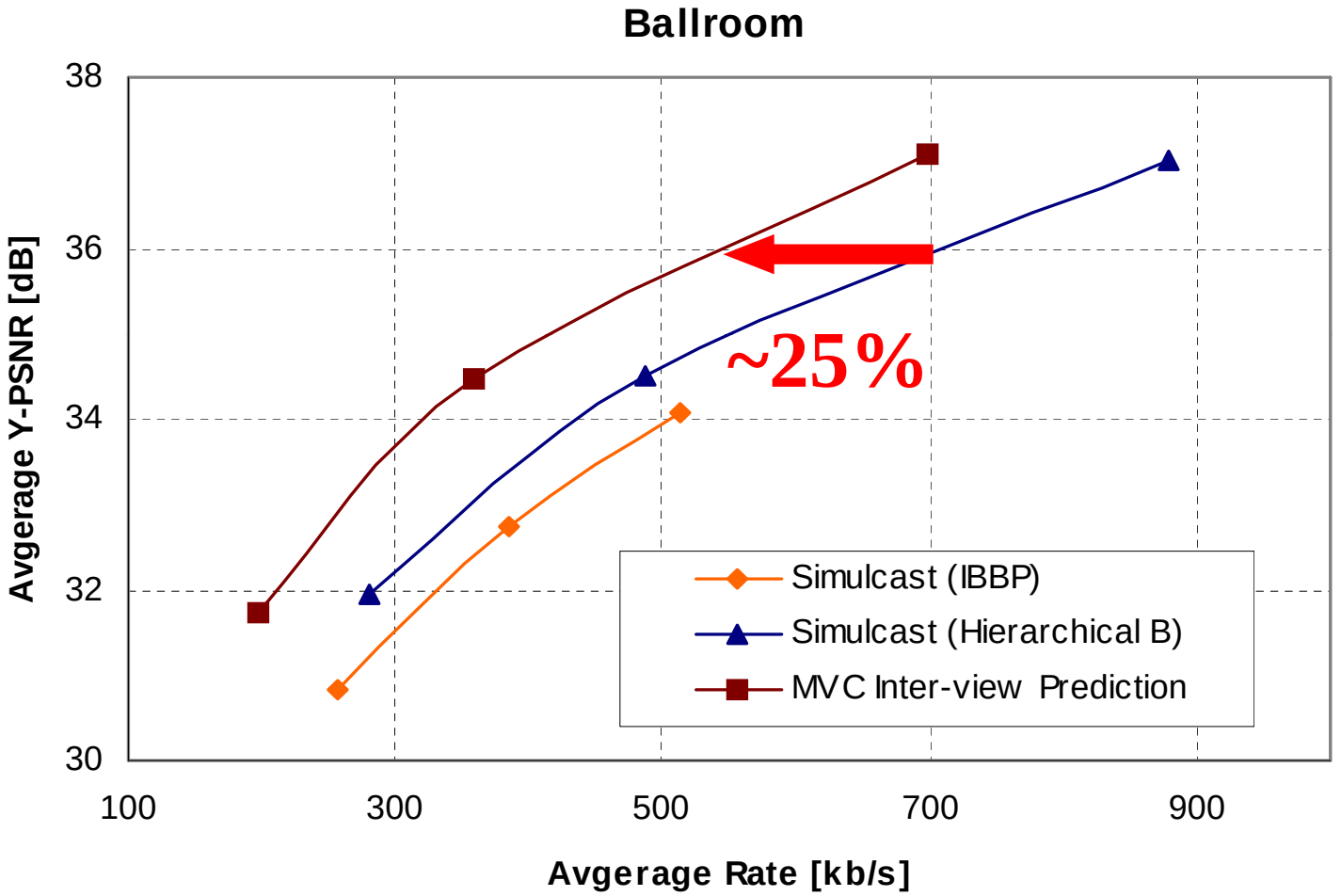


Base view independent of other views and AVC compatible

MVC Standard – Coding Efficiency

Sample comparison of simulcast vs inter-view prediction

(majority of gains due to inter-view prediction at I-picture locations)



MVC Standard – Profiles and Levels

- **MVC Profiles**
  - Profiles determine the subset of coding tools
  - **Multiview High** – finalized, part of original AVC amendment

    Supports same subset of coding tools for inter-view prediction as the existing High Profile of AVC (but **no interlaced**)
  - **Stereo High** – draft spec in 14496-10/5e Amd.1, expect to finalize by July/October 2009

    Includes **support for interlaced** and limits number of views to stereo only

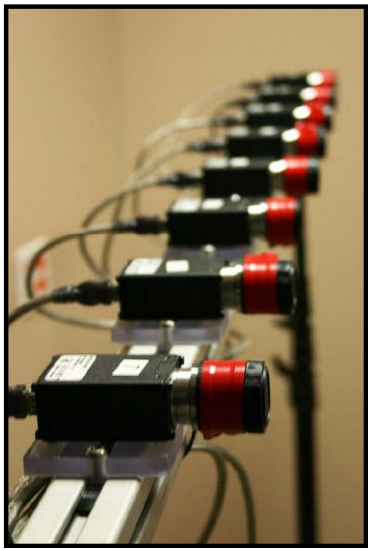
- **Level limits**
  - Levels impose constraints on resources/complexity
  - MVC repurposes fixed decoder resources of single-view AVC decoders for decoding stereo/multiview video bitstreams
  - Within a given level, tradeoff spatio-temporal resolution with number of views (e.g., specify max MBs/sec)
  - Additional constraints to enable multiple parallel decoder implementations of MVC

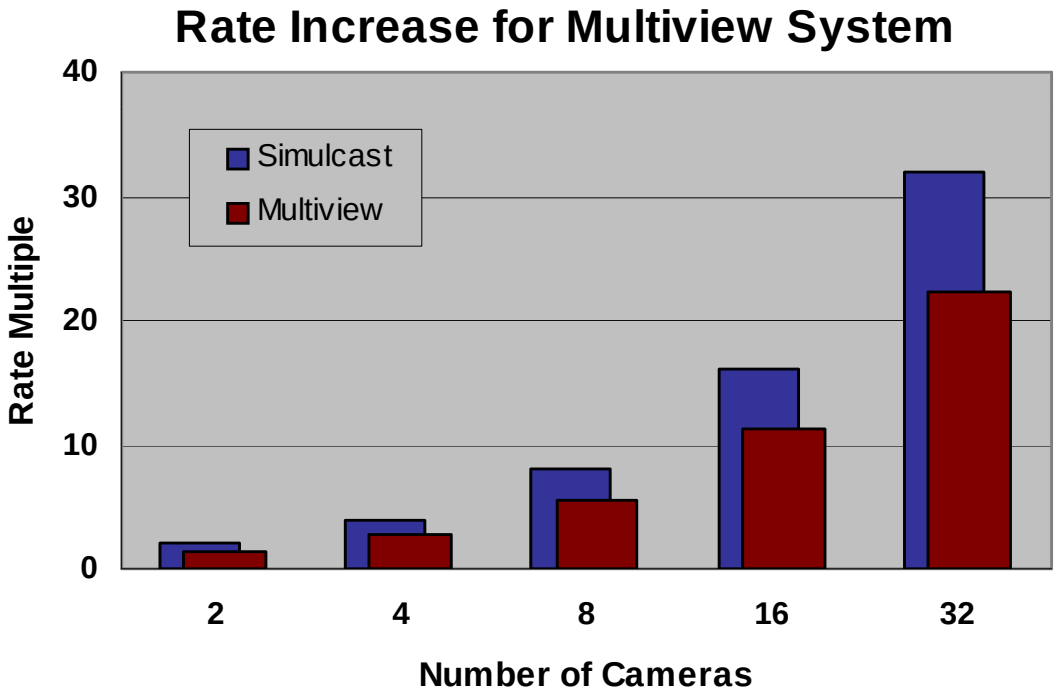
MVC Standard – Summary

- **MVC standard has recently been finalized**
  - Follow up work on conformance and transport specs underway
  - Necessary for testing interoperability and for delivery of contents to the home
- Builds on the **widely deployed AVC standard**; core encoding/decoding processes unchanged
- Offers the option to **extract a compatible 2D** representation from the 3D version

MVC Standard – Limitations/Issues

- Acquisition/production with large camera arrays is not common (and is somewhat difficult)
- Although more efficient than simulcast, rate of MVC is still proportional to the number of views
  - Varies with scene, camera arrangement, etc.





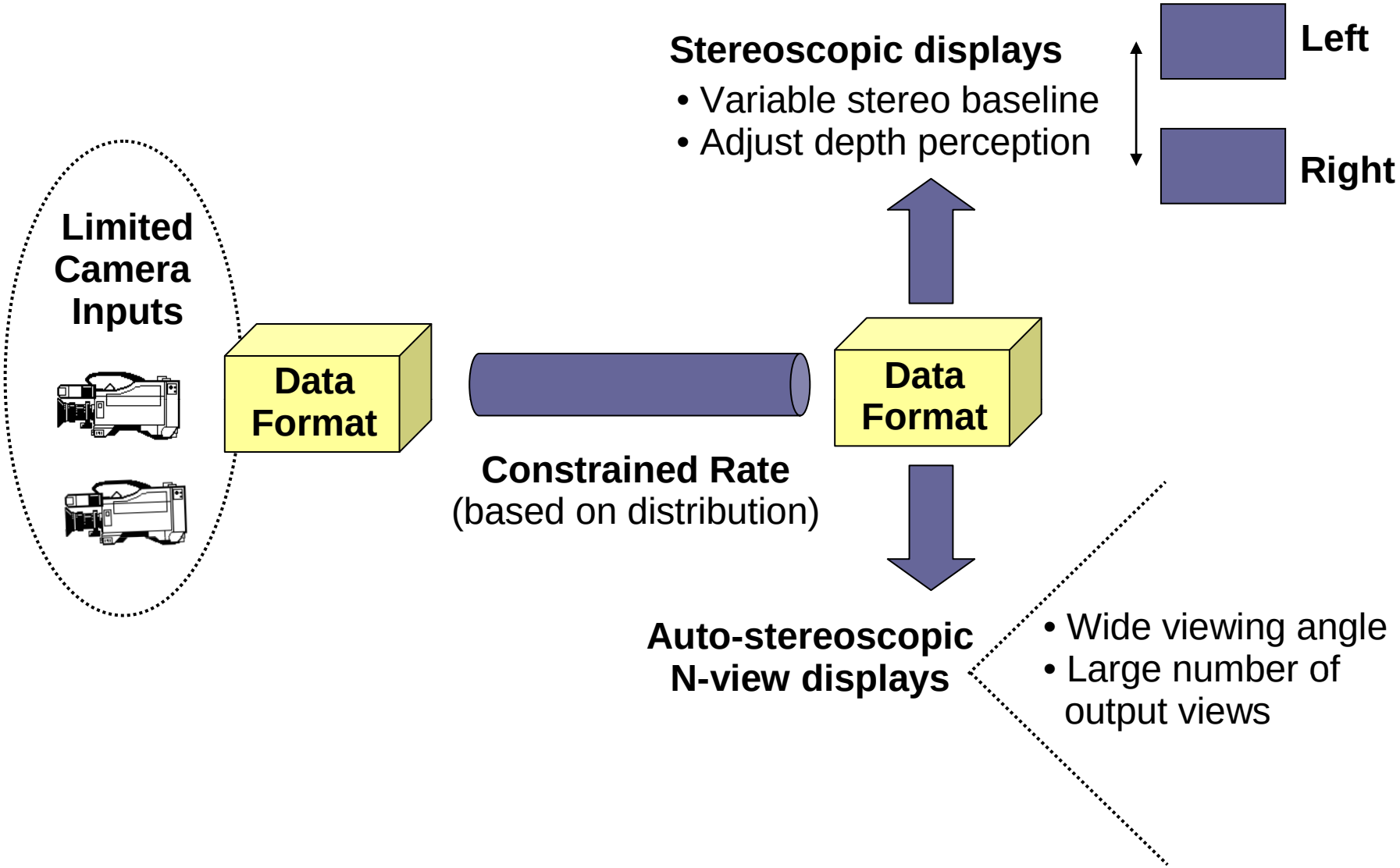
MVC Standard – Limitations/Issues

- MVC is about encoding a discrete set of multiple views
  - Goal: Highest pixel fidelity
  - Rate significantly higher than monoscopic video

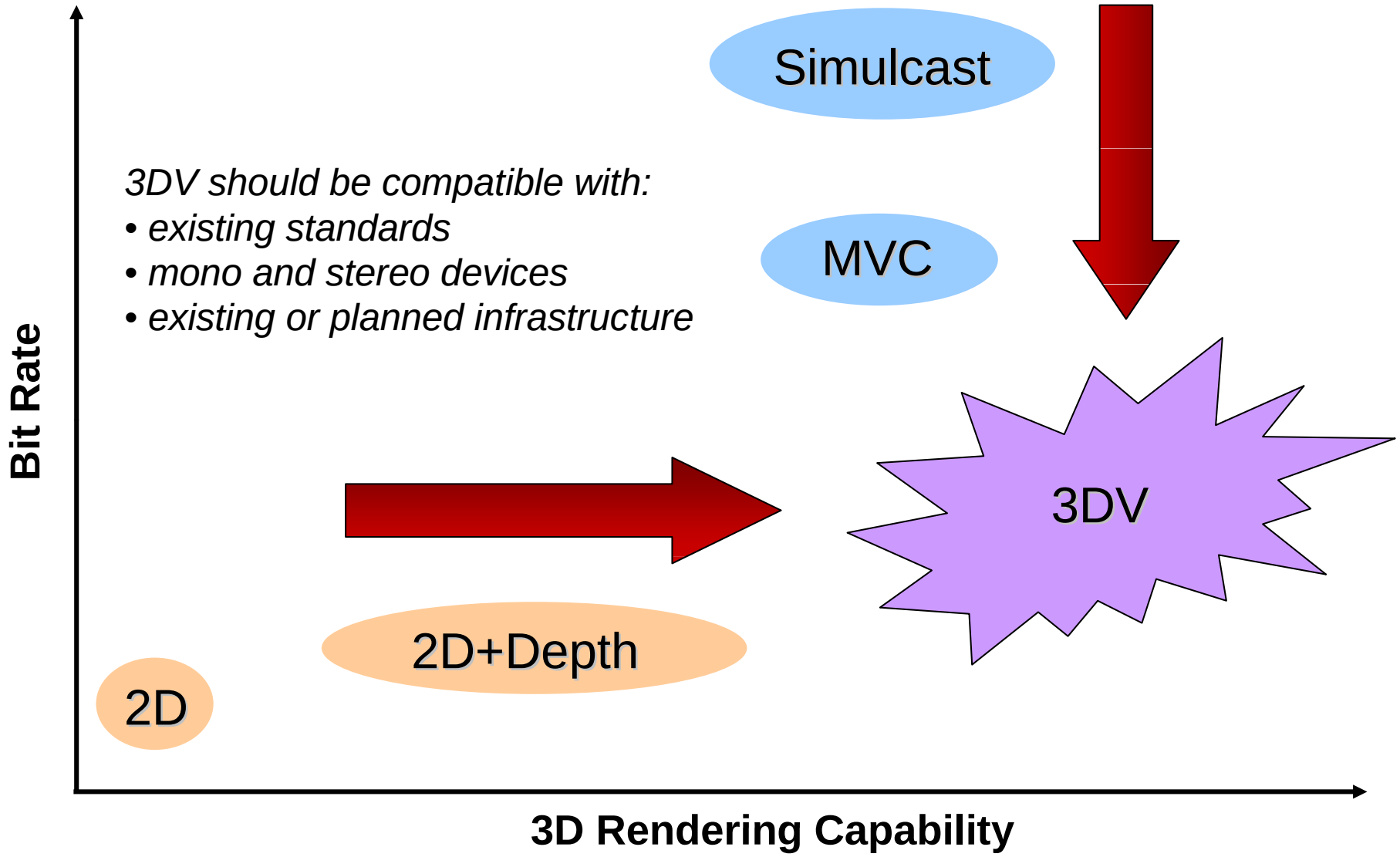
- Exploration activity in MPEG: **Free-viewpoint / 3D video** for a compressed representation and technologies allowing to generate a large number of views from a sparse view set
  - Requires depth/disparity maps representation/compression and interpolation/rendering method
  - Higher distortion may be expected (in terms of pixel fidelity, not necessarily visual quality)
  - First phase is 3D Video with expected synthesis baseline up to ±10
  -

- MPEG has already defined MPEG-C part 3 (23002-3) standard in 2006
  - Format enabling simple stereoscopic application using standard video codecs
  - Allows one video plus depth from which a second view is generated
  - Rate not significantly increased compared to monoscopic video

Anticipated 3D Video Format



Bit Rate vs 3D Rendering Capablities

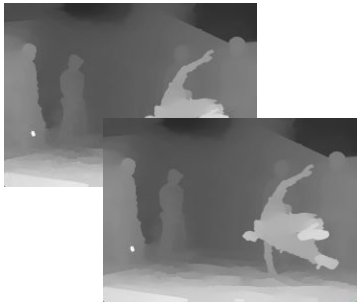


3DV should be compatible with:
- *existing standards*
- *mono and stereo devices*
- *existing or planned infrastructure*

3D Video Framework

Limited Video Inputs
(e.g., 2 or 3 views)

Depth Estimation → Video/Depth Codec → View Synthesis → Larger # Output Views



1010001010001

+

Binary Representation
& Reconstruction Process

3D Video – Observations and Remarks

- Acceptable view synthesis quality in absence of coding video and depth has been achieved
- Quality quickly deteriorates w/depth encoding
  - Fine quantization of depth causes notable artifacts, substantial increase with coarser quantization
  - Blurring artifacts with sub-sampling, might be reduced with better decimation/interpolation scheme (simple averaging used in this study)
- Better compression algorithms needed
  - Future Call for Proposals planned
- Subjective evaluation necessary
  - PSNR results not indicative of artifacts
  - New metrics could be considered

3D Video – Summary

- Main Objectives
  - Support auto-stereoscopic displays from a limited number of input views and also variable baseline for stereo processing
  - Inclusion of depth: decouple number of transmitted views with number of required views for display

- MPEG exploration underway
  - In the process of establishing suitable reference
  - Expecting to issue Call for Proposals later this year

Overall Summary

- MPEG has actively contributed compression technology for stereo and multi-view video, and is considering to take the next steps towards 3D and free-viewpoint video
- We are trying to define generic formats that are as far as possible agnostic of capturing, rendering and display processes (not easy!)
- Communication and collaboration between different bodies concerned with these matters appears necessary to avoid diversification of formats (as happened in stereo)