Mobile 3D Content

From Capture to Consumptions

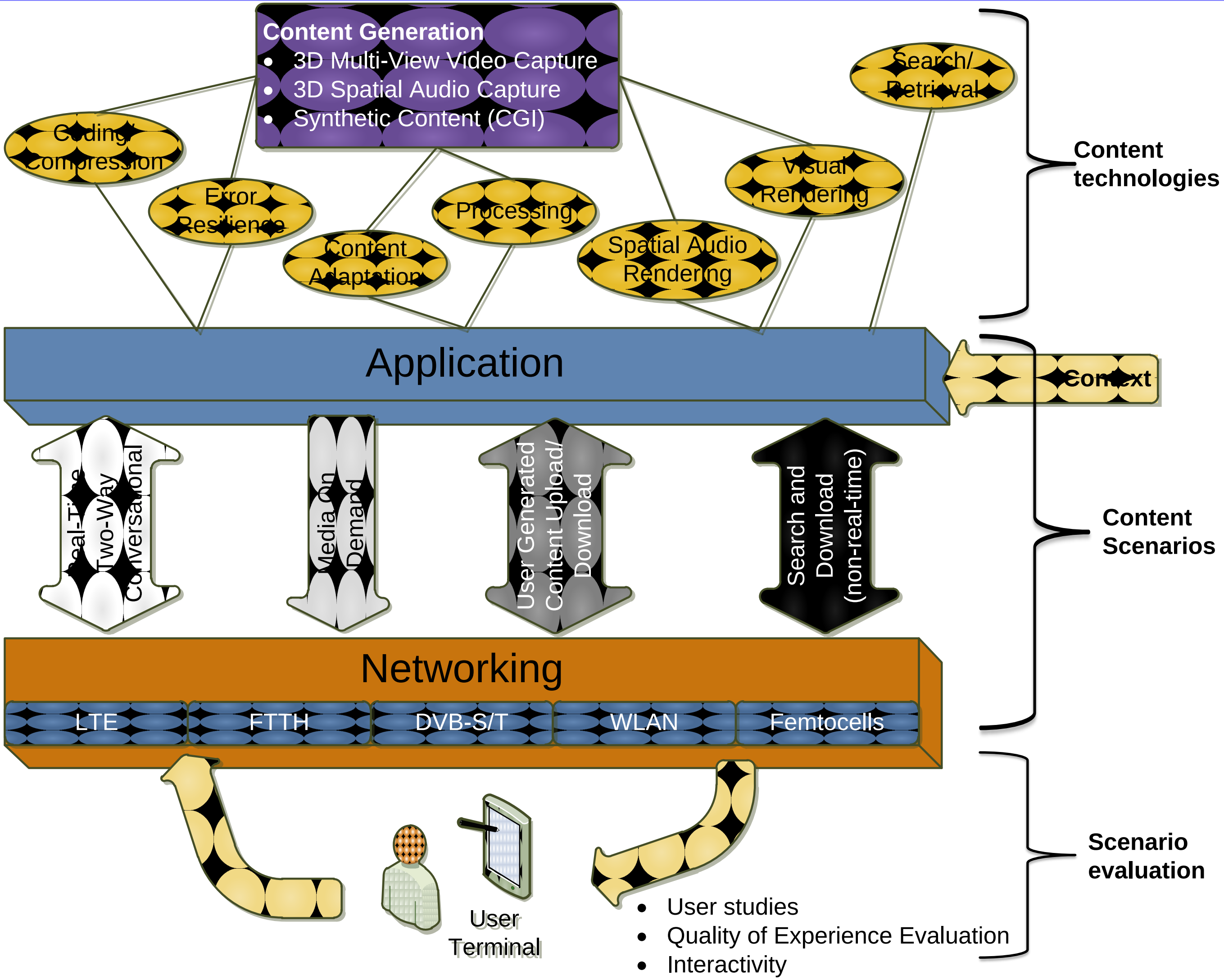
Ahmet Kondoz

I-Lab: Multimedia Communication Research

University of Surrey, UK.

Email: A.Kondoz@surrey.ac.uk

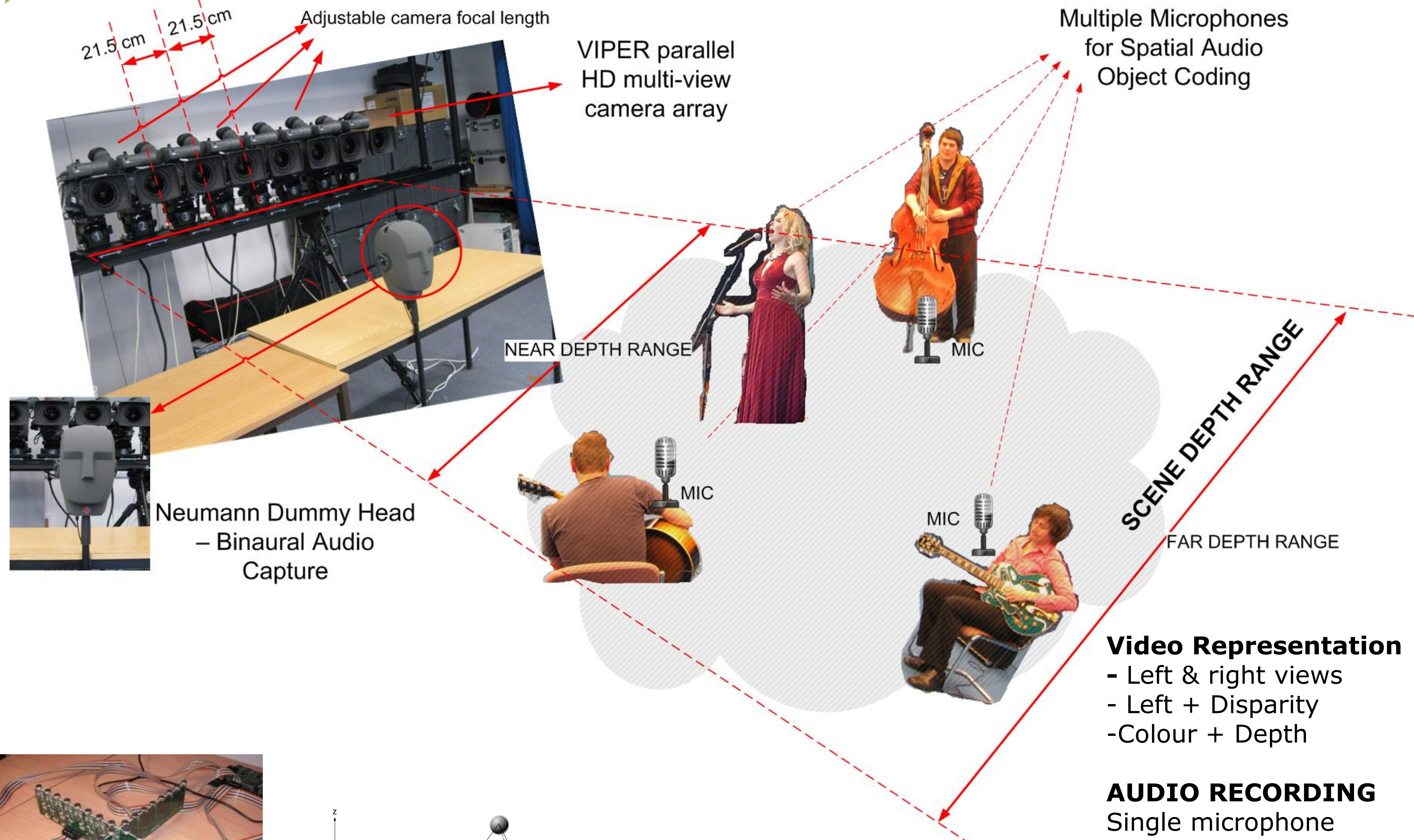
Content Stages



3D Content Capturing

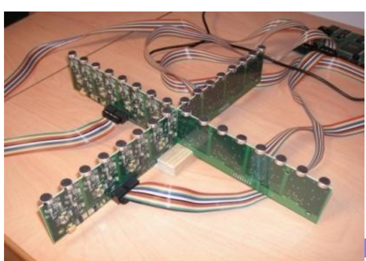
UNIVERSITY OF SURREY

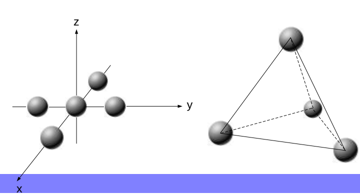
3D Audio-Visual Acquisition



Video Representation
- Left & right views
- Left + Disparity
-Colour + Depth

AUDIO RECORDING
Single microphone
Dummy head
Microphone array
Multiple microphones





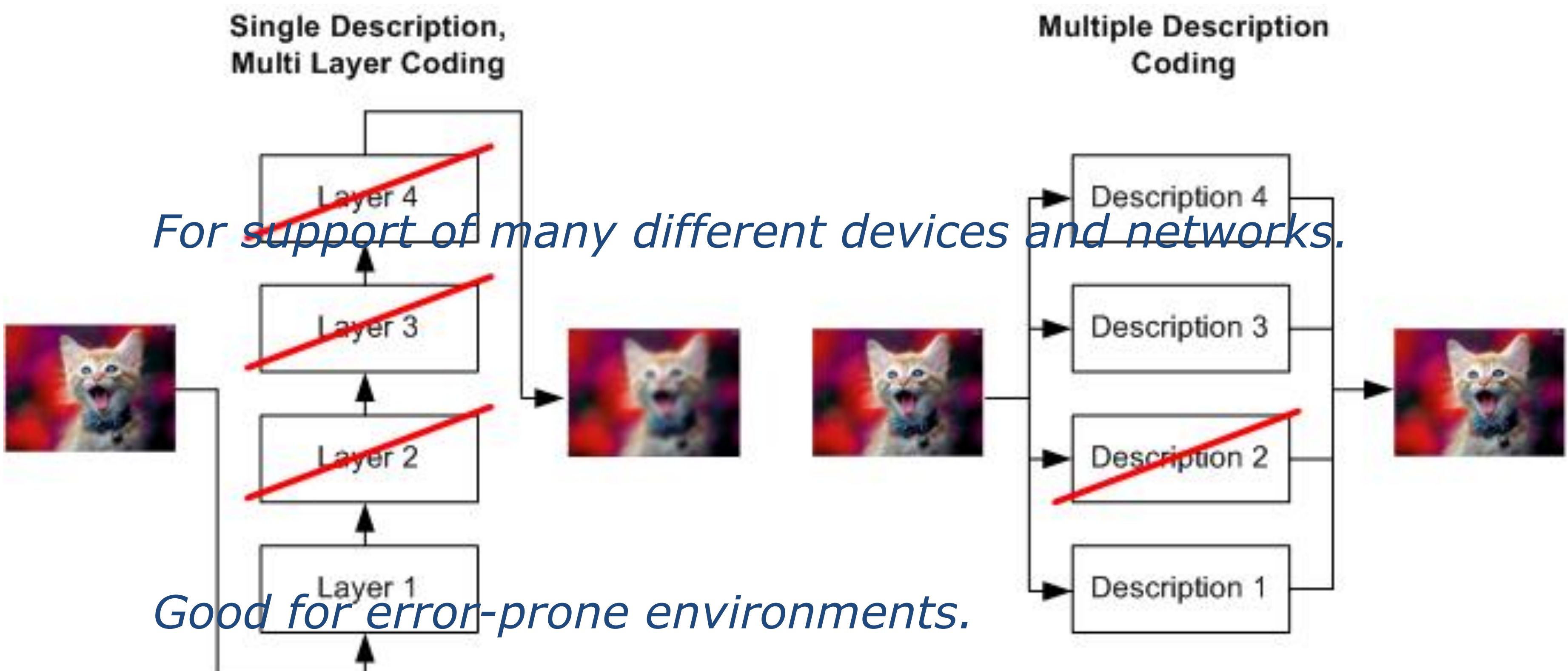
3D Content Processing

- Audio Visual coding to
  - keep the bit rate as low as possible and
  - be compatible with all possible scenarios
- Processing to
  - maintain the best audio/visual quality

Video Codec Summary

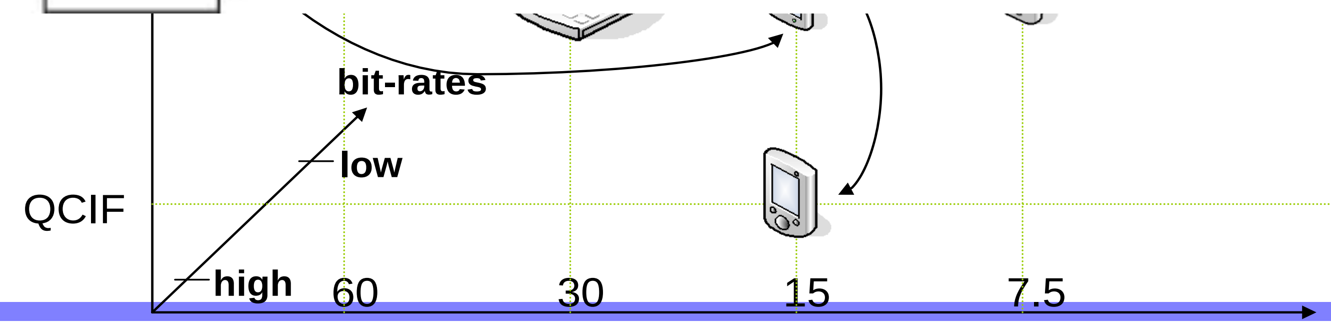
- H.264/AVC is the most popular codec
- 17 profiles for different applications
  - Each profile may have a “level” defining additional options such as resolution, bit rate etc.
- Annex G – Scalable Video Coding
- Annex H – Multi-view Video Coding
- Bit rates of 64kb/s ->300MB/s

Scalable vs. Multiple Description Video Coding



For support of many different devices and networks.

Good for error-prone environments.

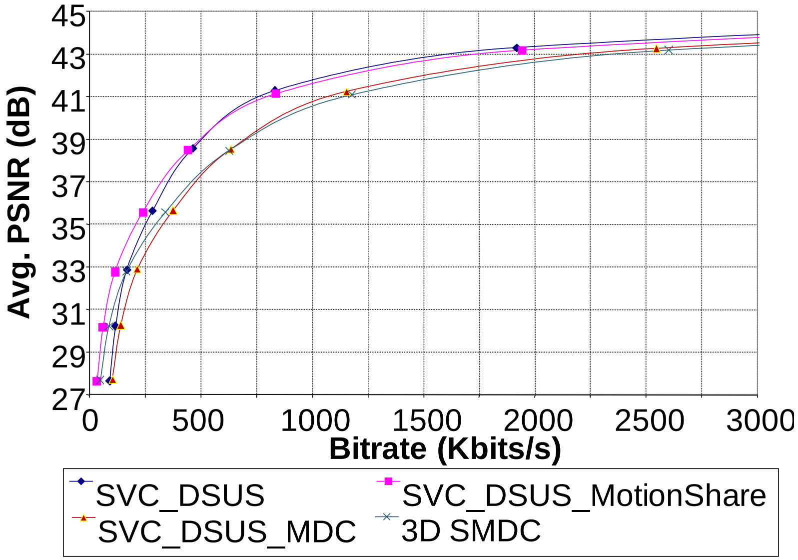


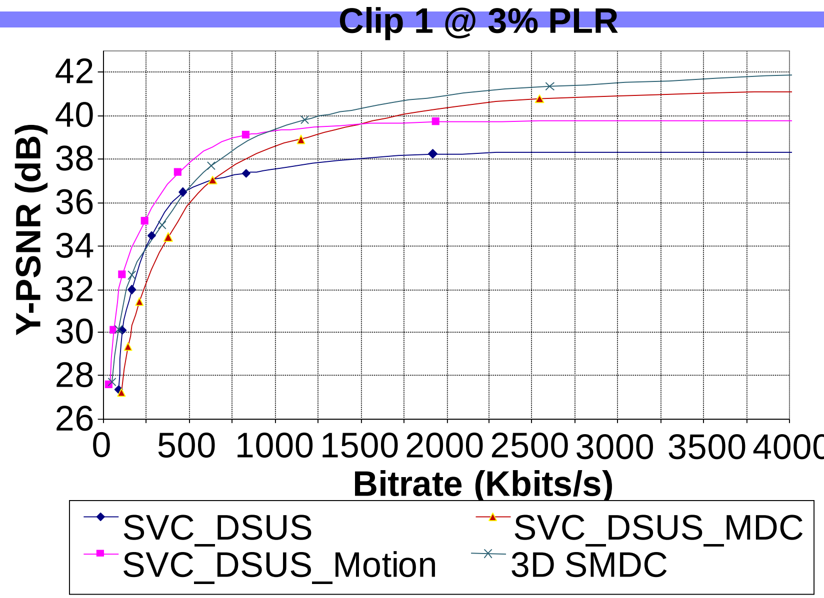
Slide source: IBBT

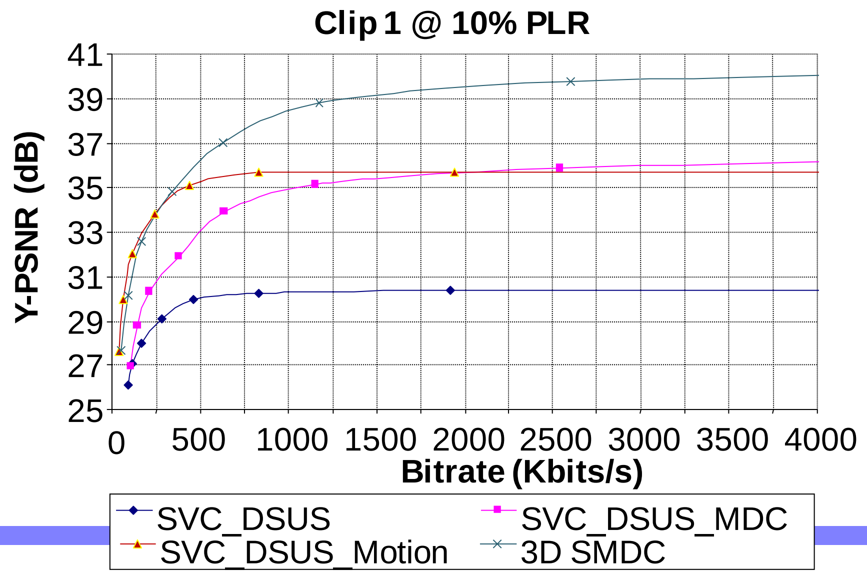
MDC Performance

SVC-DSUS (down sampling up sampling)

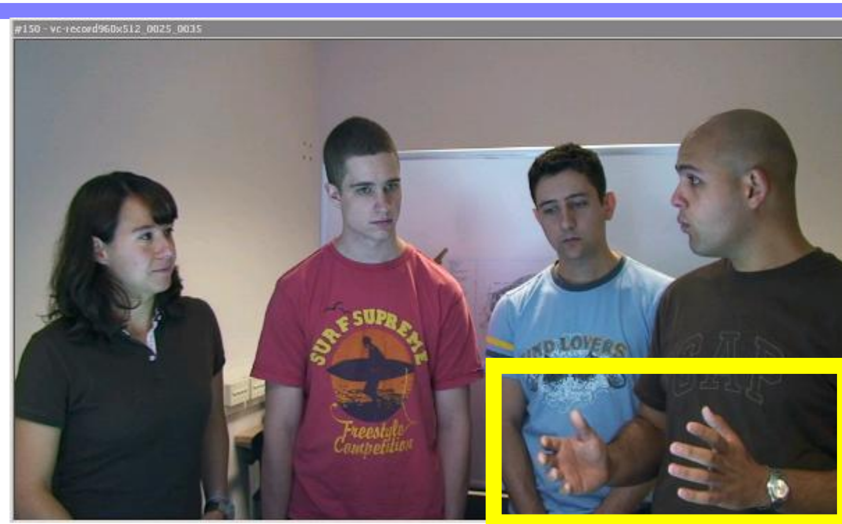
SMDC scalable MDC motion share







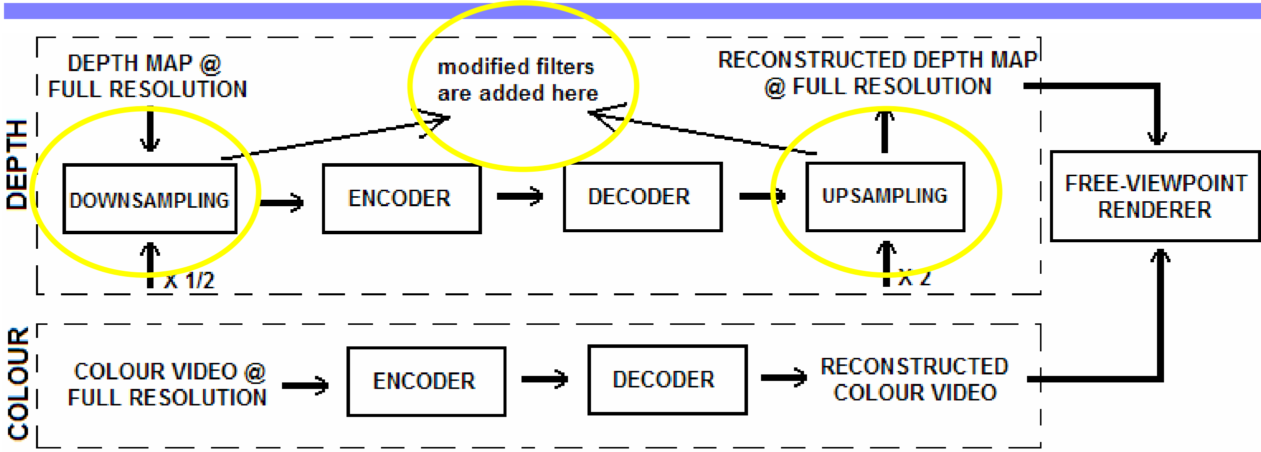
Example Frames



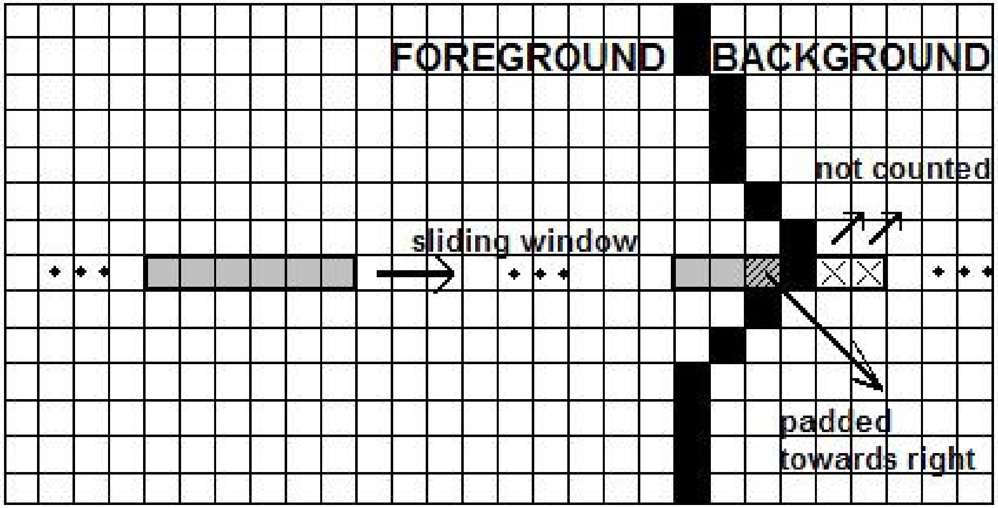
3D SVC @ 3% PLR

3D SMDC @ 3% PLR

Edge Adaptive Re-sampling for Depth Frames



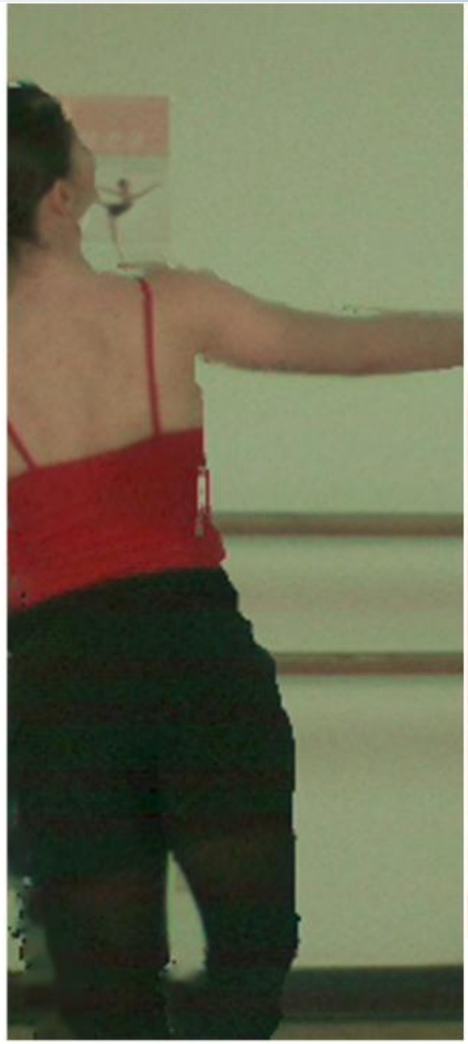
- Edge aware downsampling allows reduction in data size, while maintaining high frequency details.
- Edge aware upsampling scheme after decoding allows the conservation and better reconstruction of critical object boundaries.



- Objective is to reduce depth map coding overhead.

Edge Adaptive Re-sampling for Depth Frames

with edge adaptive upsampling



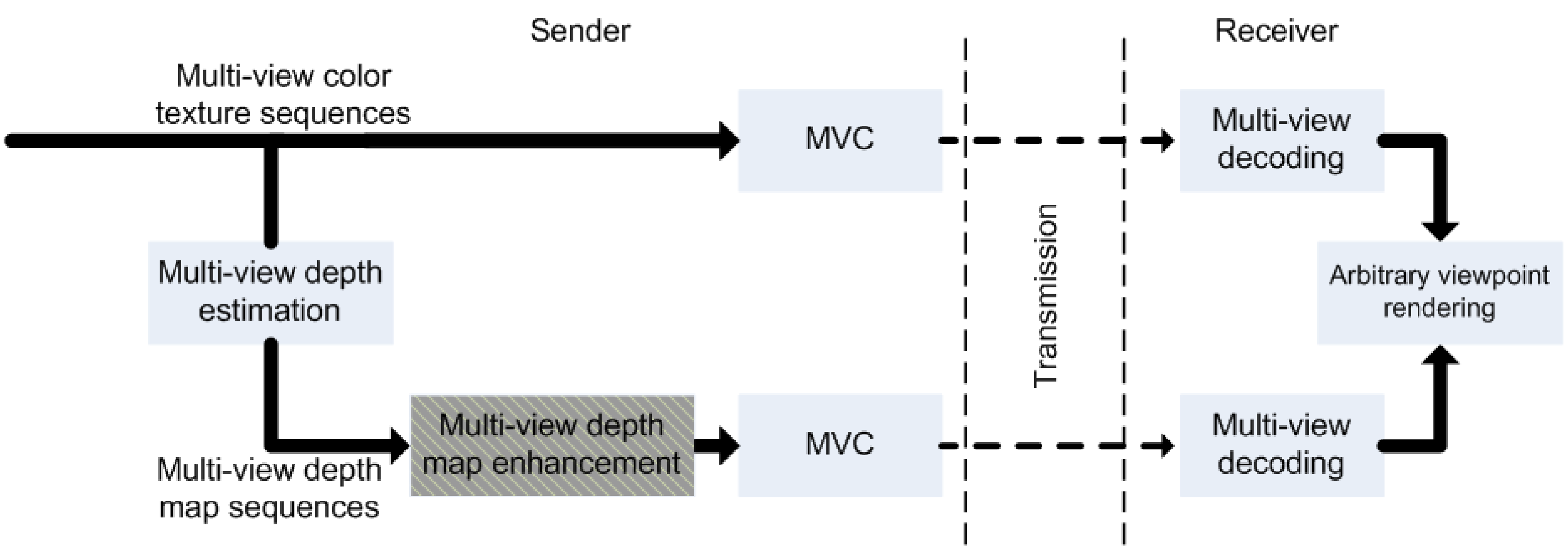
without edge adaptive upsampling

with edge adaptive upsampling

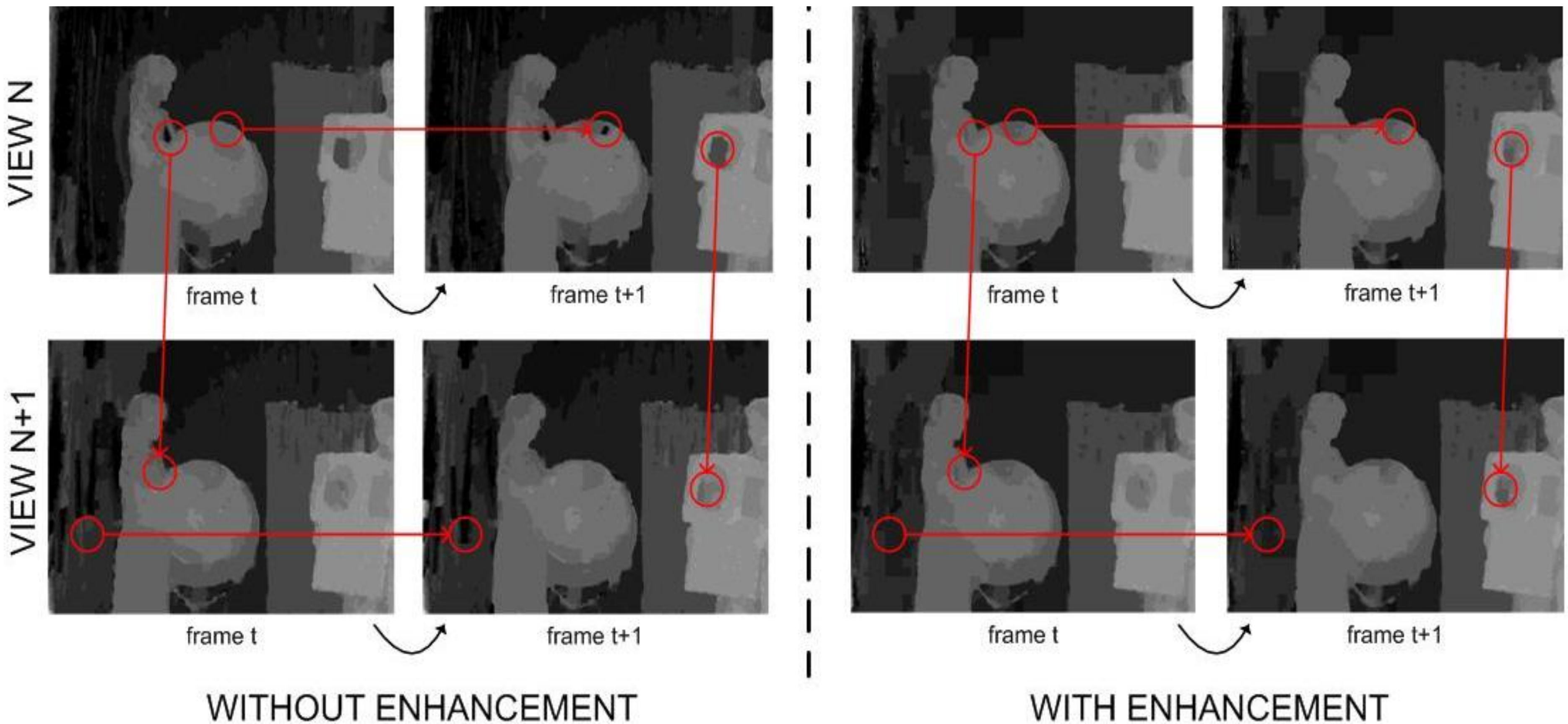
without edge adaptive upsampling

Multi-view Depth Map Enhancement via Adaptive Median Filtering

- Multi-dimensional median filter is applied on multi-view depth maps
  - ✓ Increased consistency along time axis within same view
  - ✓ Increased inter-view depth coherence
  - ✓ **Improved coding and rendering performance**

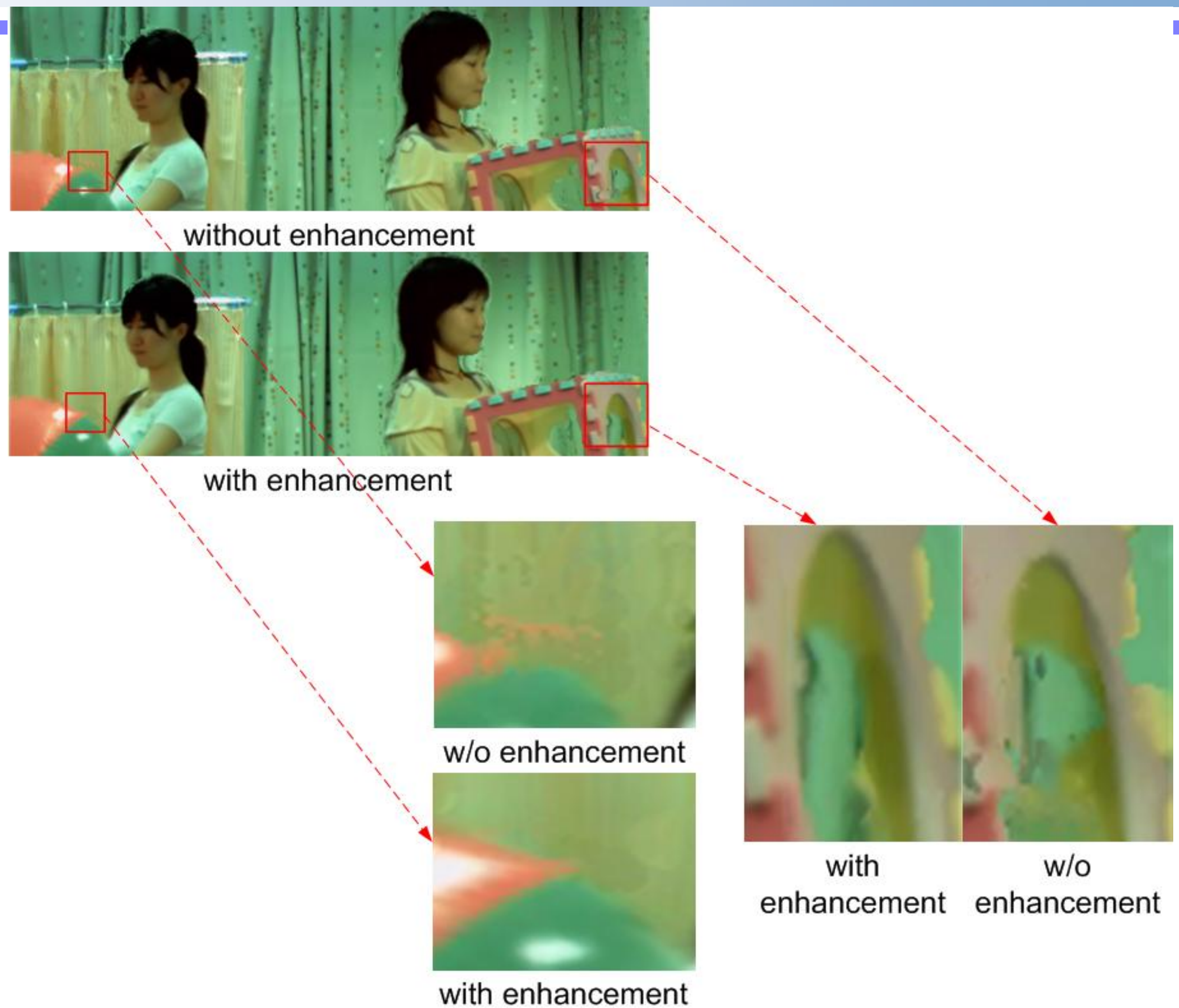


Multi-view Depth Map Enhancement via Adaptive Median Filtering

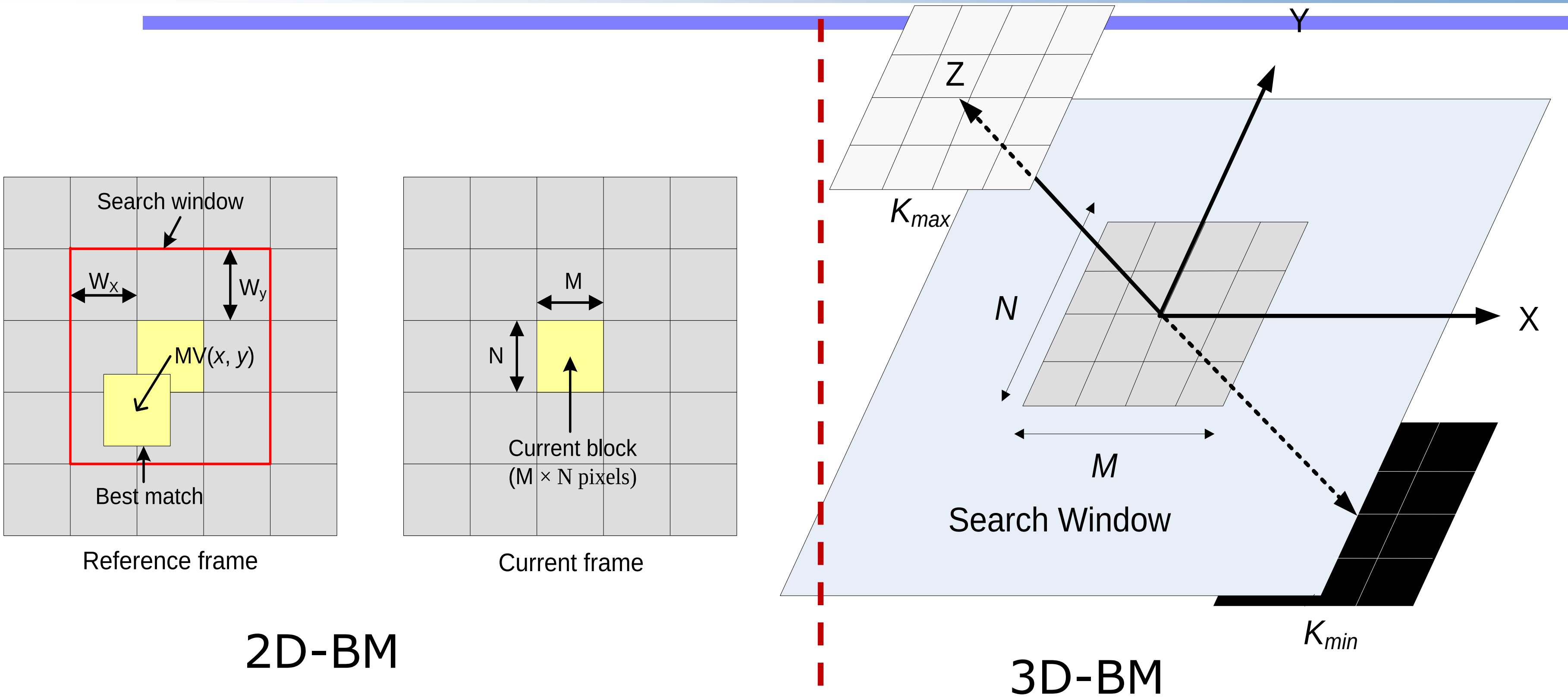


Multi-view Depth Map Enhancement via Adaptive Median Filtering

Example FVV rendering results:

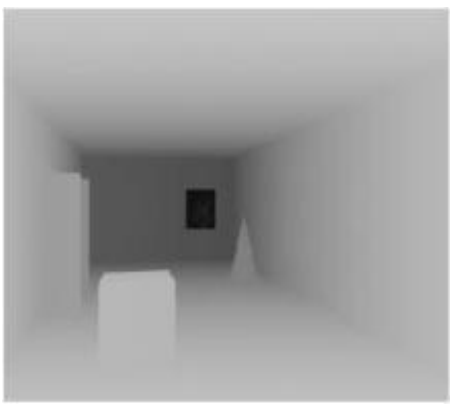


3D Block Matching (3D-BM) for Depth Video Coding

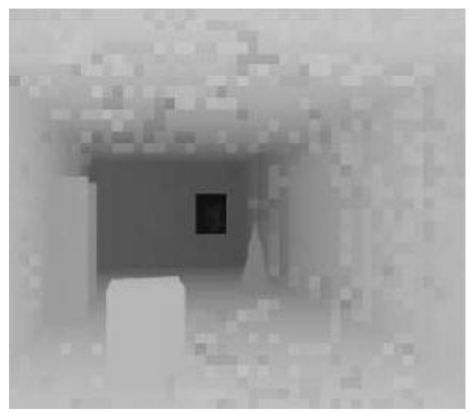


- Pixel values indicate a relative distance between objects and a camera.
- Pixel values change if objects move in depth direction.

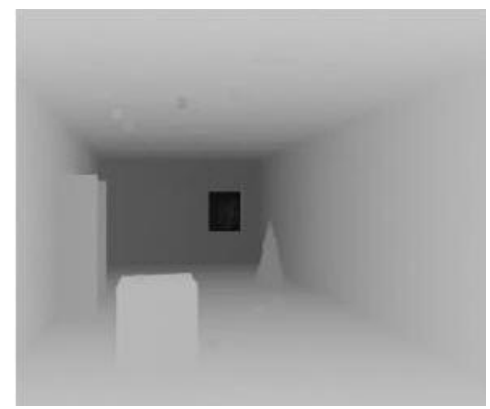
Comparison of Motion-Predicted Signals



Original frame



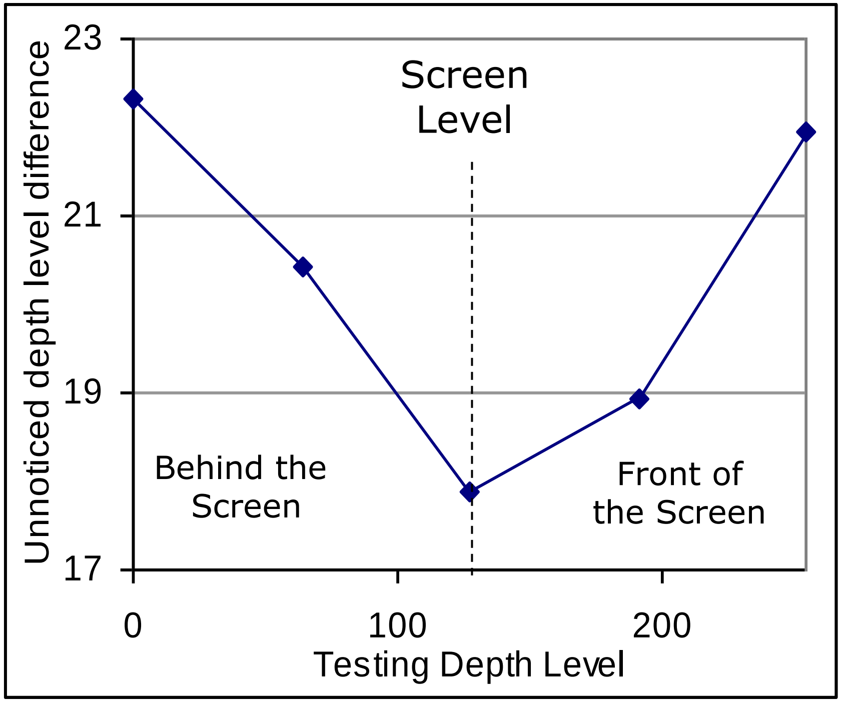
Predicted signal from 2DBM



Predicted signal from 3D-BM

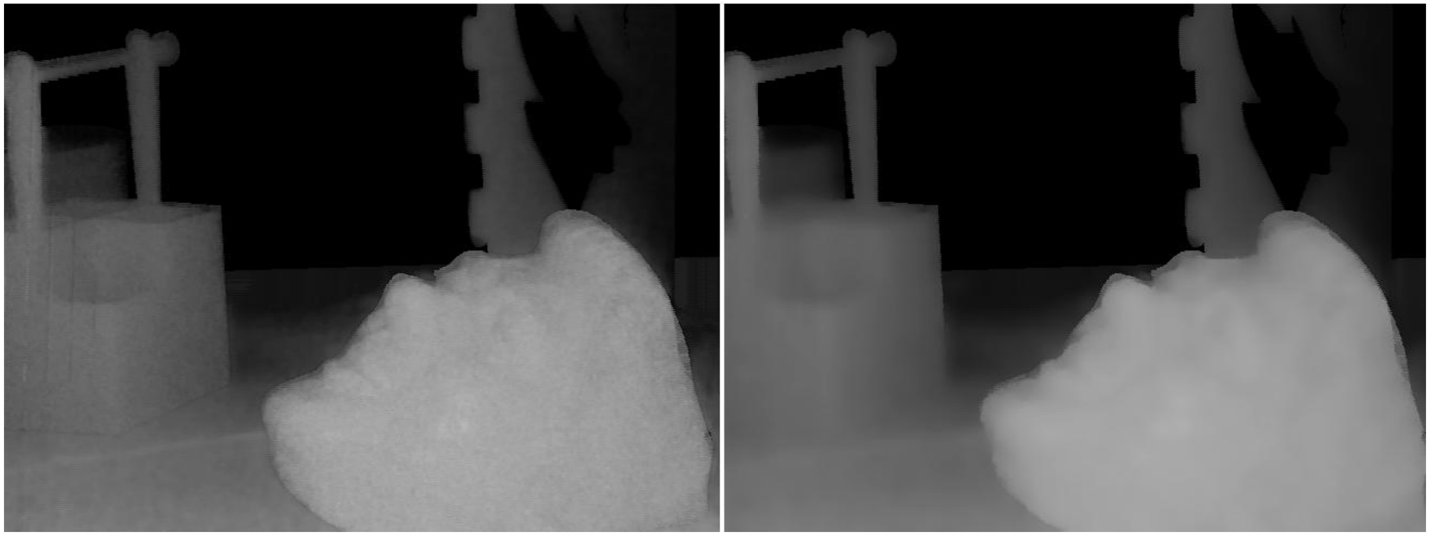
Modelling of the perceived depth sensitivity in 3D video

- Humans can not usually perceive sufficiently small depth changes in a scene
- Experimentally derived a Just Noticeable Difference in Depth (JNDD) model to apply to a stereoscopic 3D video display system



Average of unnoticed depth level difference at various depth levels

Depth map pre-processing based on the JNDD model

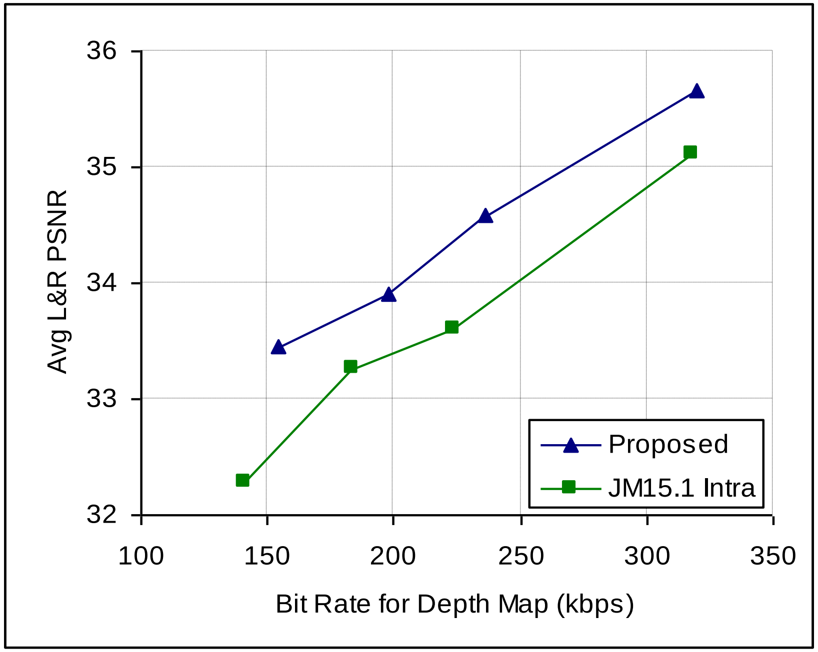


(a) 'Orbi' Original unprocessed depth map

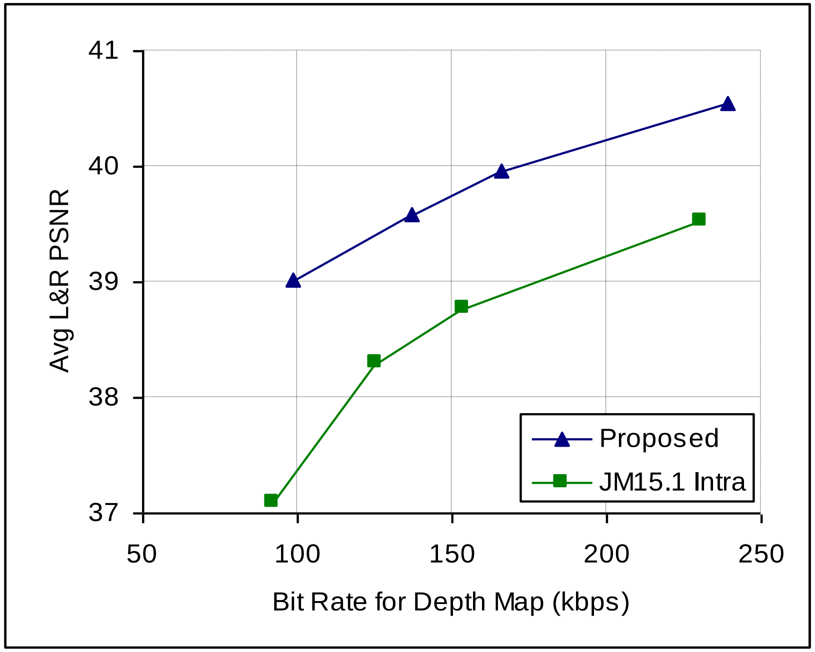
(b) pre-processed with the method based on JNDD Model

- depth map sequences are pre-processed to suppress the depth details that are not perceivable by the viewers
- This will minimise the irregularities in the rendering process that arise due to optical noise
- Bit rate for depth map coding can be reduced up to 70% (sequence dependent)

Results of depth image based rendering (DIBR)



Ballet



Breakdancers

- Y-axis = Average PSNR of rendered Left and Right views with the depth map
- X-axis = Bit rate required to encode depth map in kbps

Audio Processing

UNIVERSITY OF SURREY

AMR Codec

Narrow Band AMR (@8kHz)

Mode	Bitrate (kb/s)	Channel
AMR_12.20	12.20	FR
AMR_10.20	10.20	FR
AMR_7.95	7.95	FR/HR
AMR_7.40	7.40	FR/HR
AMR_6.70	6.70	FR/HR
AMR_5.90	5.90	FR/HR
AMR_5.15	5.15	FR/HR
AMR_4.75	4.75	FR/HR
AMR_SID	1.80	FR/HR

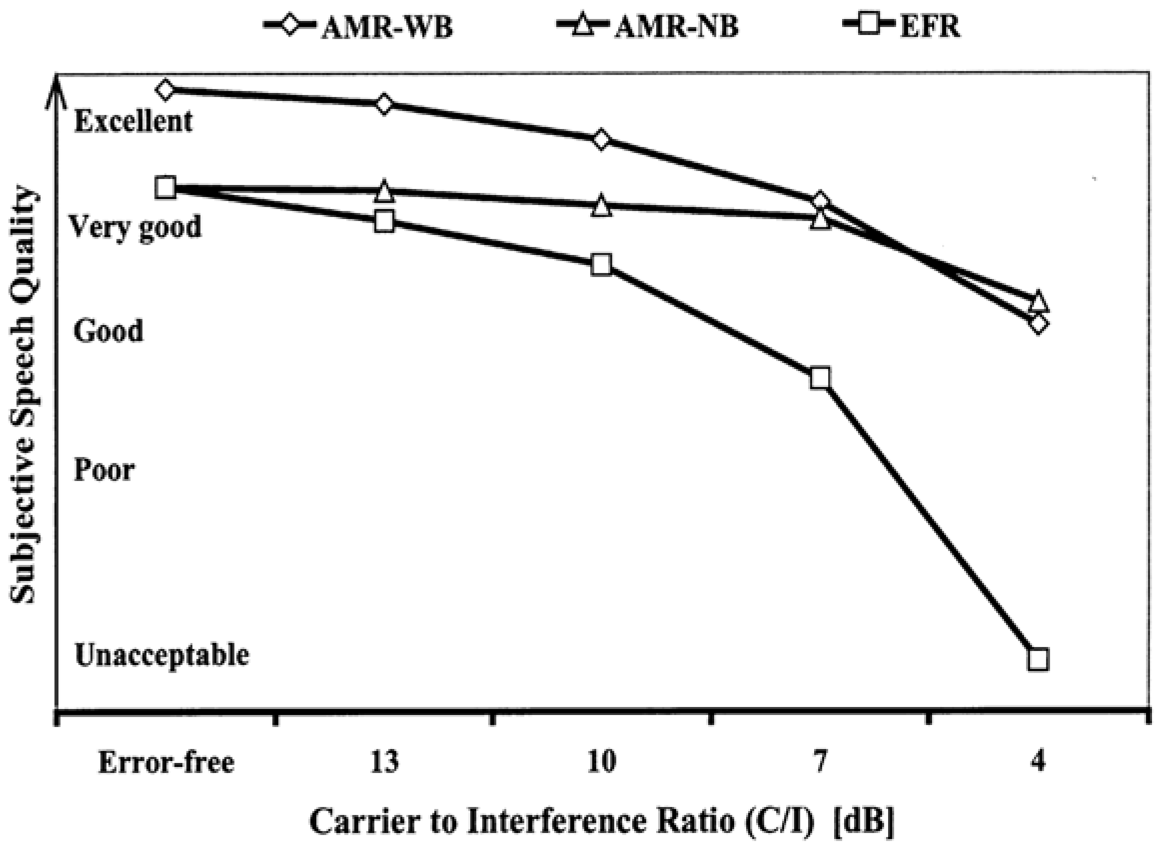
Wide Band AMR (@16kHz)

Bitrate (kb/s)
23.85
23.05
19.85
18.25
15.85
14.25
12.65
8.85
6.60

AMR WB+ 5.2-48kb/s @ 44.1kHz

AMR Performance in Mobile Channels

- AMR-NB(0.3-3.4 kHz) samples at 8kHz
- AMR-WB(0.3-7.4 kHz) sampled at 16kHz
- AMR-WB+(0.3-20 kHz) samples at 44.1 kHz

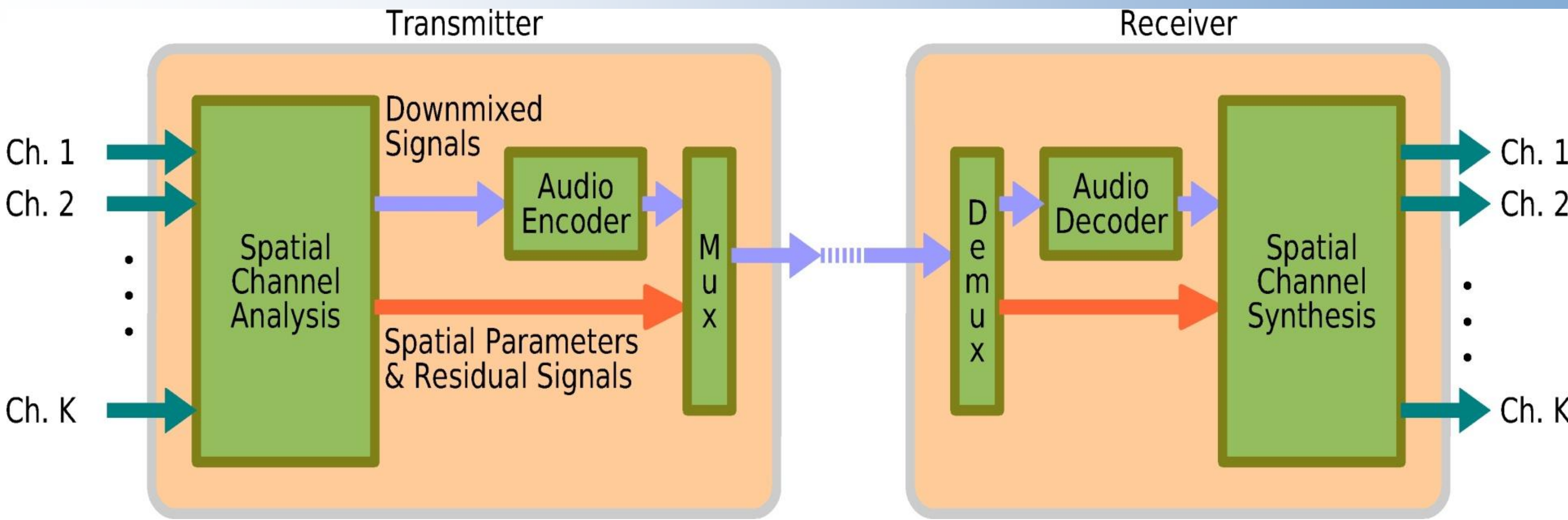


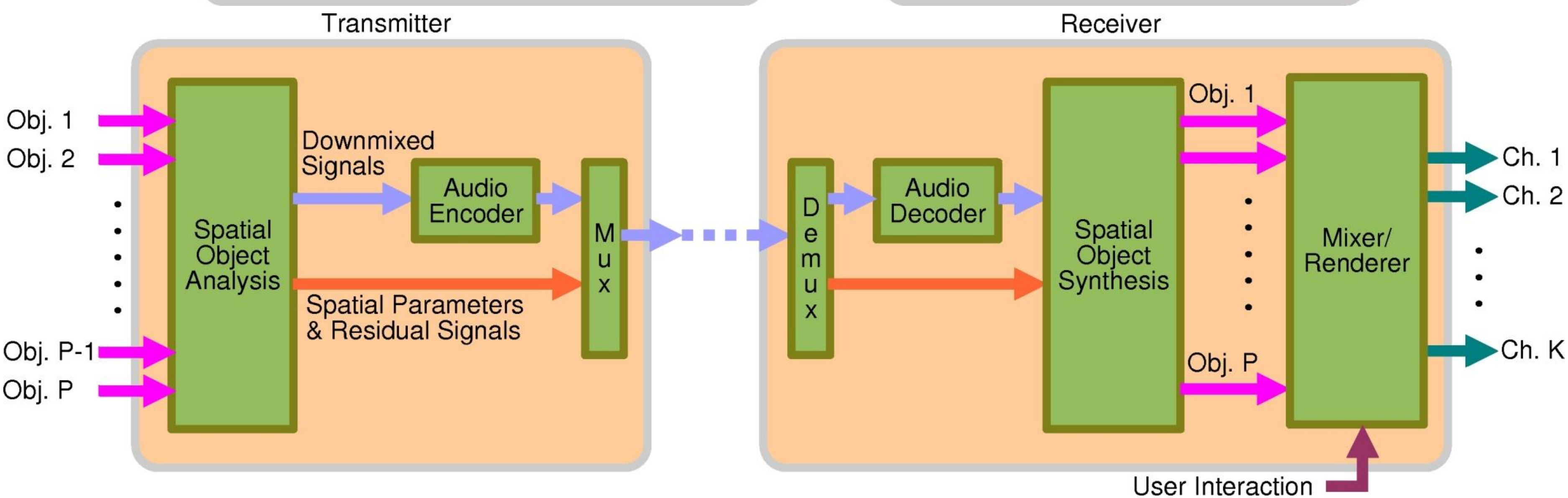
3D Audio Coding

Technology	Description	Example
Spatial Audio Coding	Spatial parameters are extracted from multichannel audio signals and then transmitted as side information along with mono or stereo downmixed signals	MP3 Surround, MPEG Surround
Spatial Audio Object Coding	Audio scene is represented by several audio object rather than in multichannel audio signals. Spatial parameters are then extracted from object signals and transmitted along with mono or stereo downmixed signals	MPEG-4 SAOC

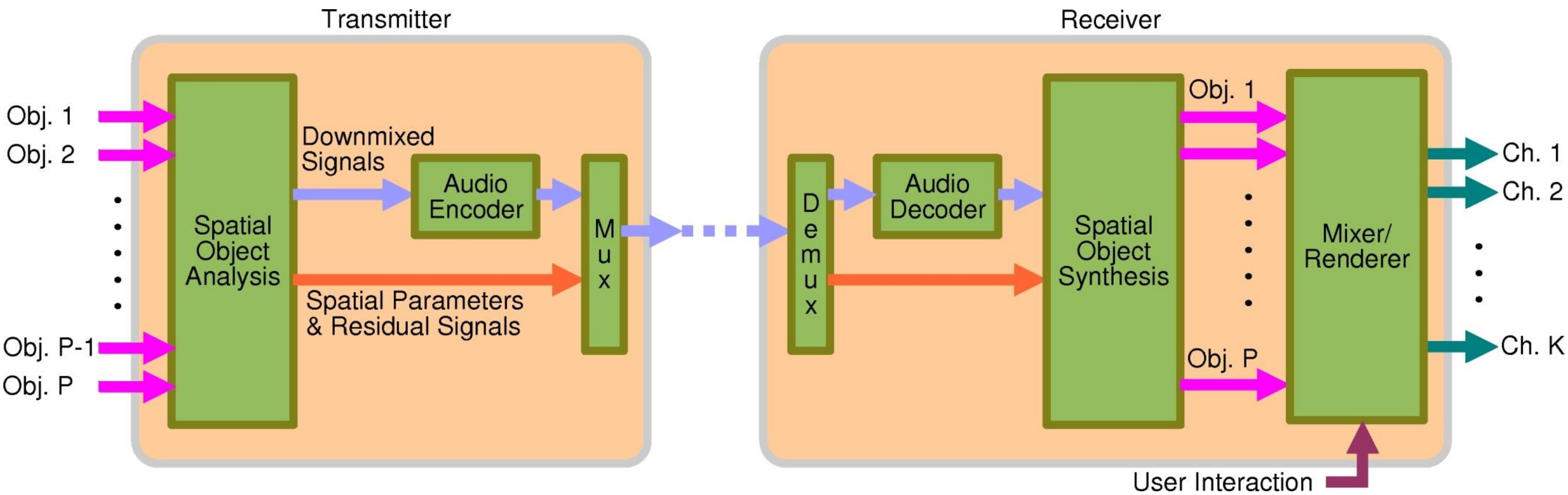
Typical rates 48-256 kb/s

Spatial Audio and Spatial Audio Object Coding





Spatial Audio Object Coding



- Listen to each player separately
- Make an instrument quieter or louder
- Adaptive 3D reproduction independent of the rendering system

Content Adaptation

Content Adaptation Concept

- Growing heterogeneity in mobile media
  - Device capabilities
  - Access network characteristics
  - Content representation formats
  - Natural environment of users
  - User preferences
  - ...

Content adaptation is the process of transforming a media stream to another media stream to meet diverse resource constraints and user preferences while optimising the overall usability of the multimedia content

3D Content Adaptation Specifics for Mobile Applications

Mobile device specific adaptations

- Small display sizes
- Lightweight – Limited processing capability

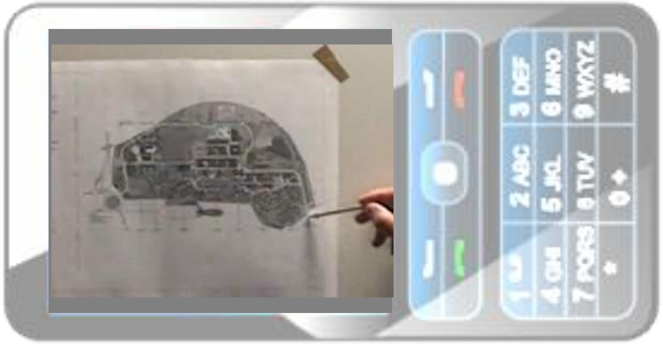
Mobile/wireless channel specific adaptations

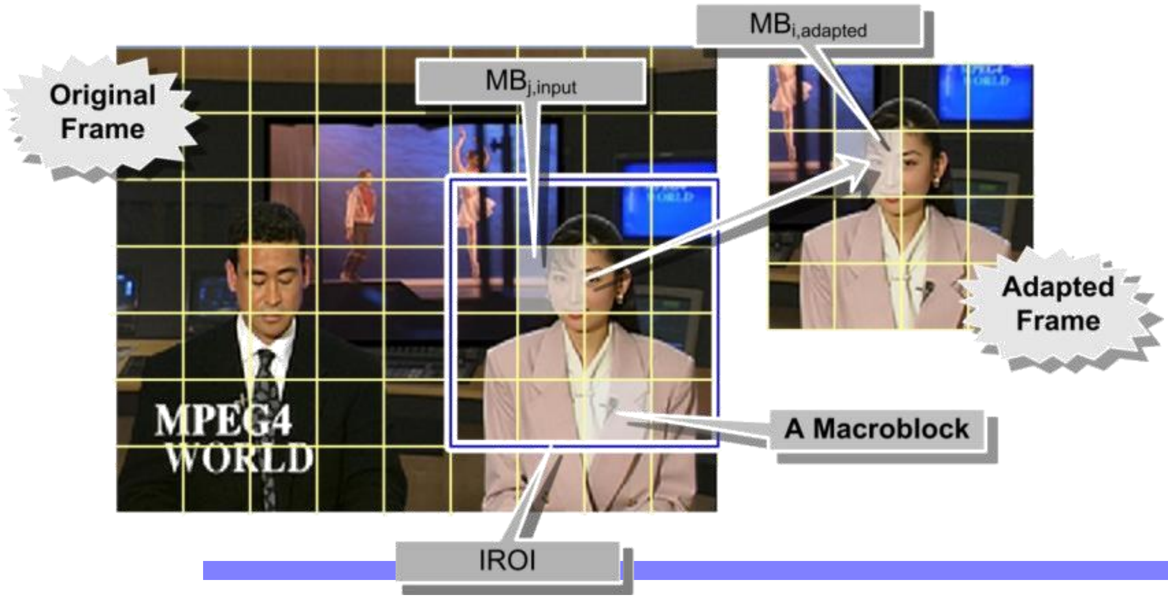
- Narrow bandwidth
- Error prone channels

- **Adaptation options**
  - Depth /colour-texture spatial, temporal, quality scaling
  - Illumination options
  - Conversion of 3D content to 2D
  - View dropping during multi-view content access
  - Cropping and scaling
  - Error robustness provision
  - Prioritised levels for scalable layers in 3D content
  - .

Adaptation via ROI Cropping & Scaling of Non-Scalable Video for Mobile Devices

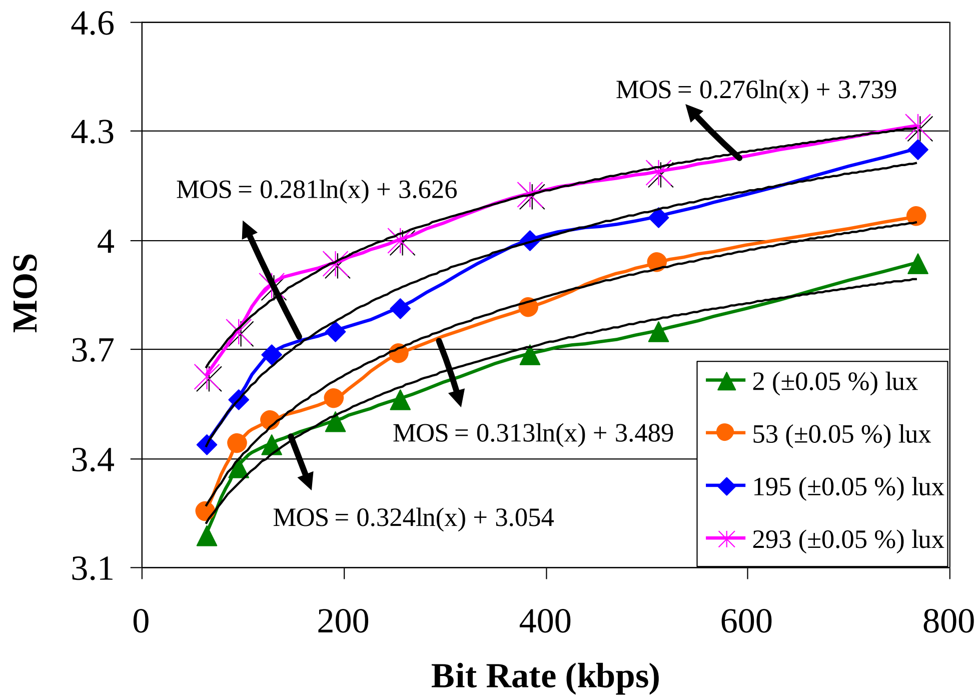
- Reducing the spatial resolution is a straightforward mechanism to adapt high-resolution (e.g., HD) video to mobile devices fitted with small and low-resolution displays
- Not an ideal adaptation method for videos with important attention areas
- Hence, cropping of video, so that enlarged attention area can be viewed on the small display, is a better adaptation solution
- What is the point of transmitting the whole picture – should this be done at source? Feed back – delay etc.

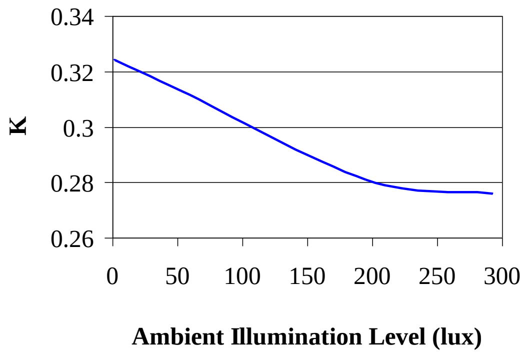


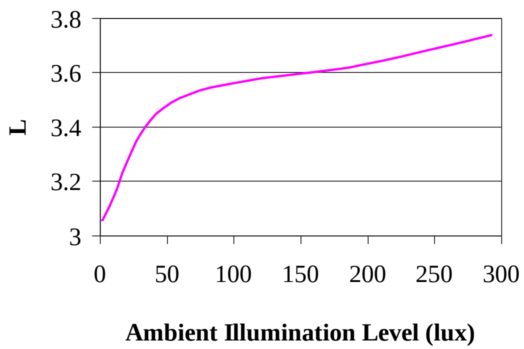


Perceptual Video Quality under Different Ambient Illumination Levels

- An example set of subjective assessment results for the Football sequence:







- Generic function patterns of the curves:

$$MOS = K\ln(B) + L$$

constant bit rate constant

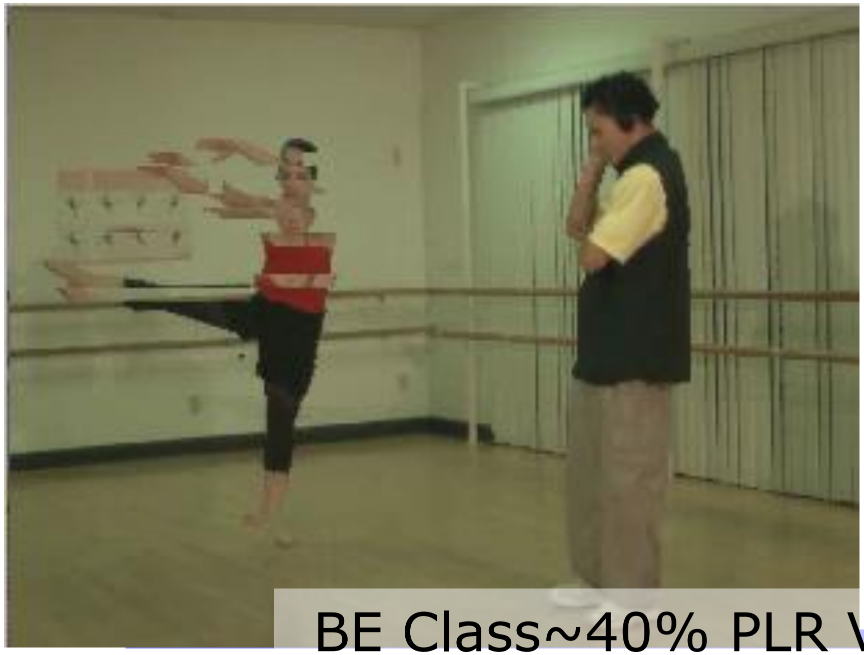
Error Resilience Adaptation for Wireless Communications (802.11e WLAN)

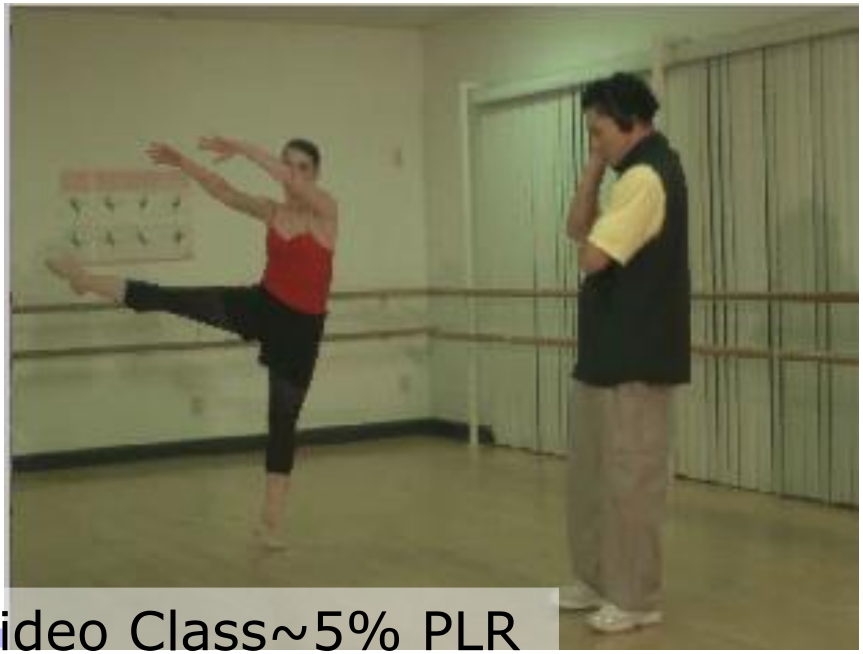
Simple Prioritised

- Foreground
  - Tx with Video Access Class
- Background+depth
  - Tx with Best Effort Class

Quality Based Prioritised

- Estimate quality based on channel PLR
  - Use quality estimate to allocate video packets to different traffic classes





BE Class~40% PLR Video Class~5% PLR

Other Content Adaptations…

- Select the best temporal, spatial and quality options based on a generic utility function
- 3D content adaptation using video/audio attention models
- Audio assisted video adaptation and vice versa

3D Content Rendering

Mobile 3D Rendering Challenges

- Screen size and processing power on the Mobiles
- Un-controlled Environment
- Variation of the Content Quality due to
  - Channel Noise
  - Bandwidth Variations

3D- Mobiles

Nokia

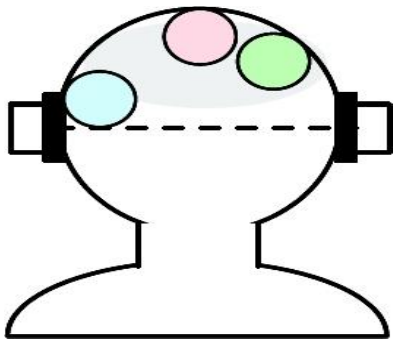
The latest Sharp 3.4” parallax barrier 3D touch screen LCD

Hitachi 3.1”

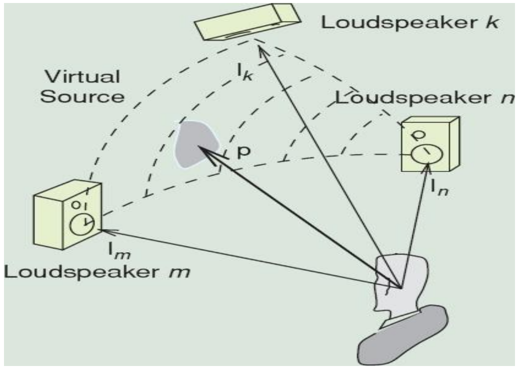
Nintendo to launch
3D portable game console

3D Audio Rendering

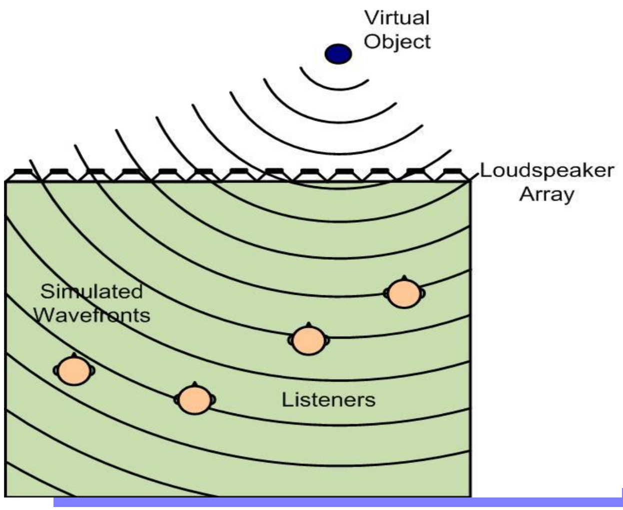
Binaural Rendering Using Headphone
(C. Faller, Spatial Audio, 2005)



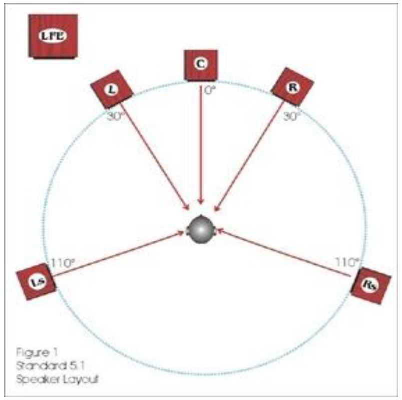
Multi Loudspeaker using Panning Techniques such as VBAP and Ambisonics
(V. Pulkki and M. Karjalainen, 2008)



Multi loudpseaker using Wave Field Synthesis (WFS)



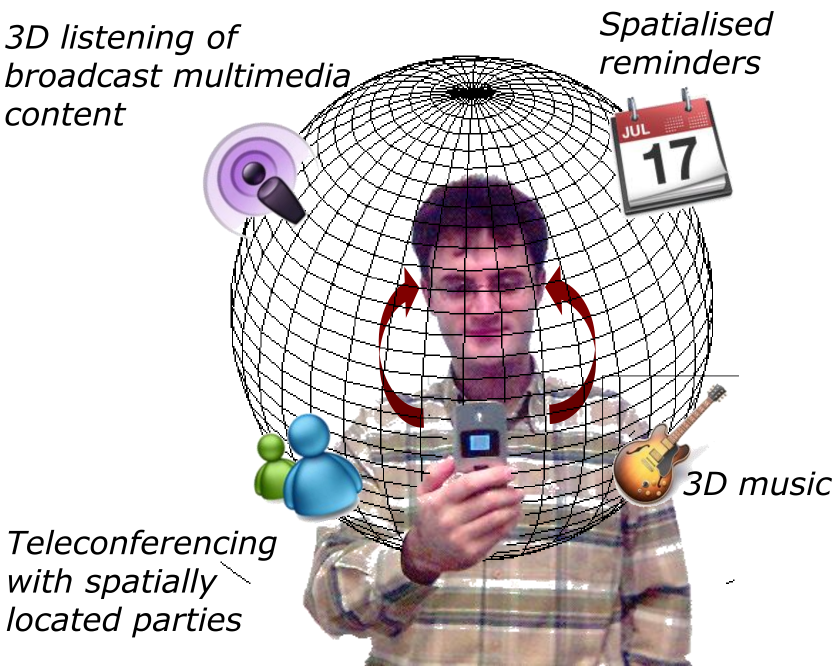
Standard multichannel audio reproduction such as 5.1, 7.1



(www.microsoft.com/windows/windowsmedia/howto/articles/surroundsoundcodecs.aspx)

Directional Audio Synthesis

- Unwanted sounds can be filtered out based on their directions using a spatial filter:



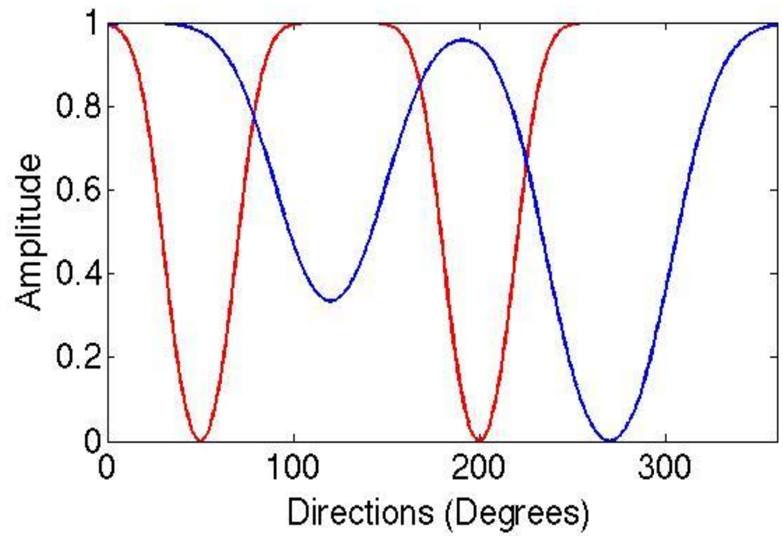


Fig.: *Two spatial filter examples based on von Mises functions for suppression of sounds at 50º and 200º with a beamwidth of 40º (red) and at 120º and 270º with different suppression levels with a beamwidth of 70º (blue).*

Quality of Experience

From QoS to QoE

We need:

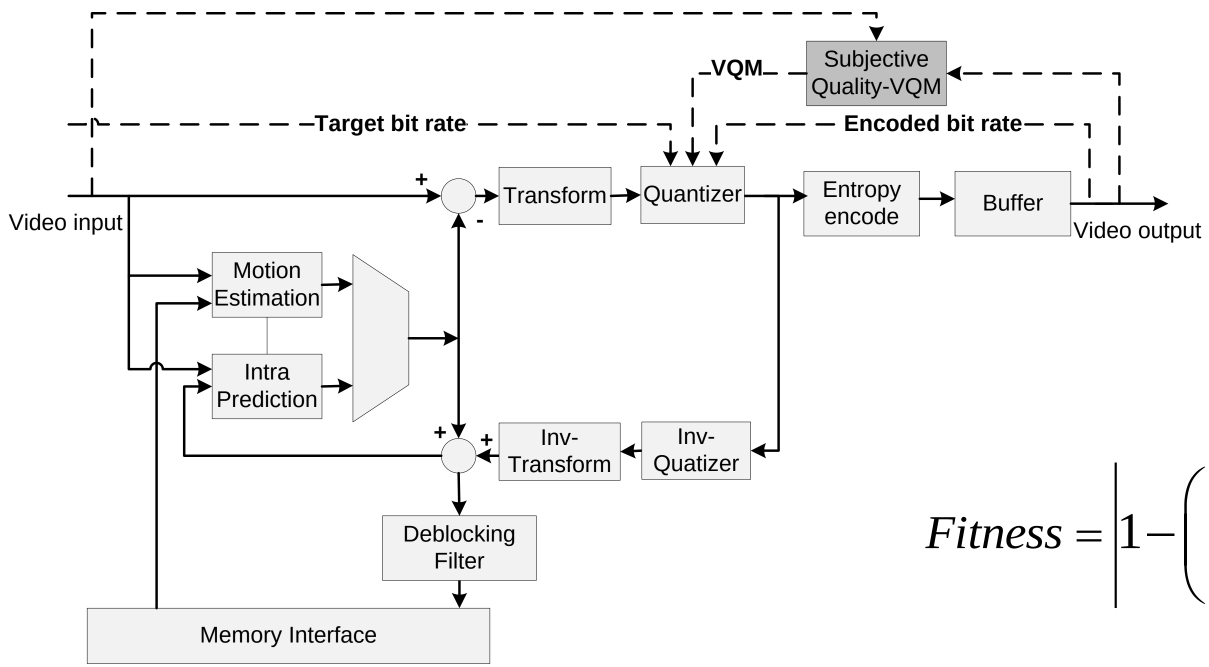
*‘**QoE models** that will allow automatic system configuration and optimization for the end-to-end multimedia (3D) delivery chain, which will **enhance the expectations** of the users by:*

*Enabling the best possible sensation, perception and emotion for **each task**’*

Examples: Conference call, on-line work/business, entertainment, socialisation …

Audio and Video importance in the content (horror movie, sports)

Perceptual Rate Control…

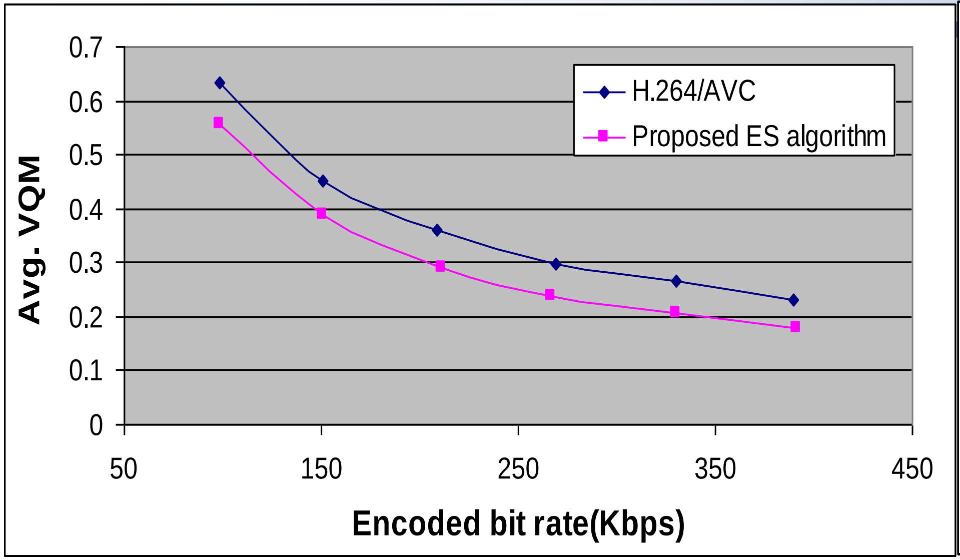


$$Fitness = \left|1 - \left(\frac{encoded\ bitrate}{target\ bitrate}\right)\right| + \lambda[VQM]$$

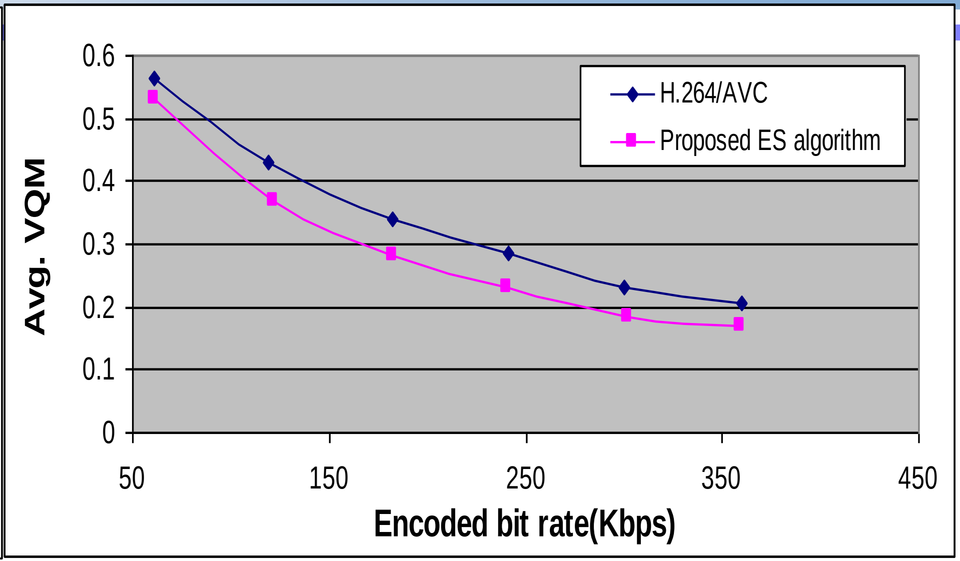
$$VQM = [VQM_{Left} + VQM_{Right}]/2$$

- Problem has been defined using
  - Perceptual quality of the output video (VQM)
  - Target bit rate
  - Actual encoded bit rate

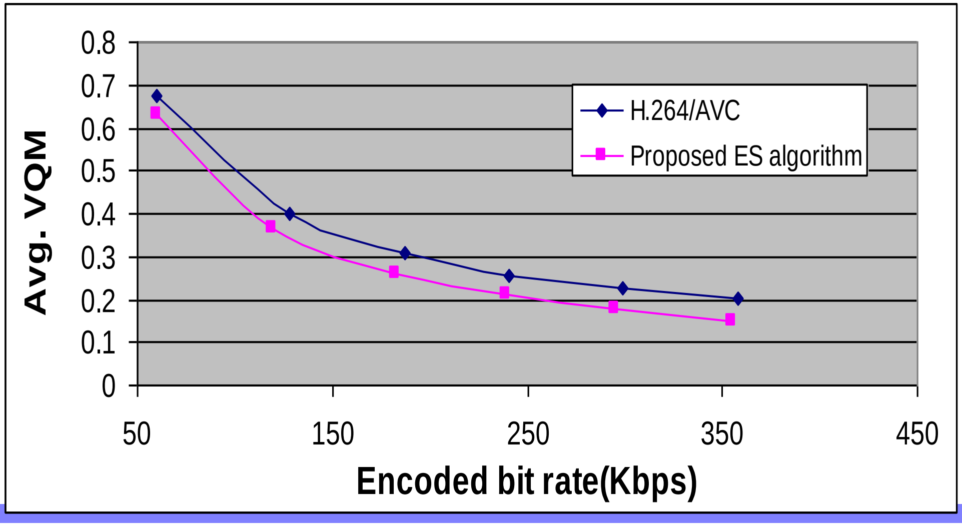
Indicative Results…



Room sequence



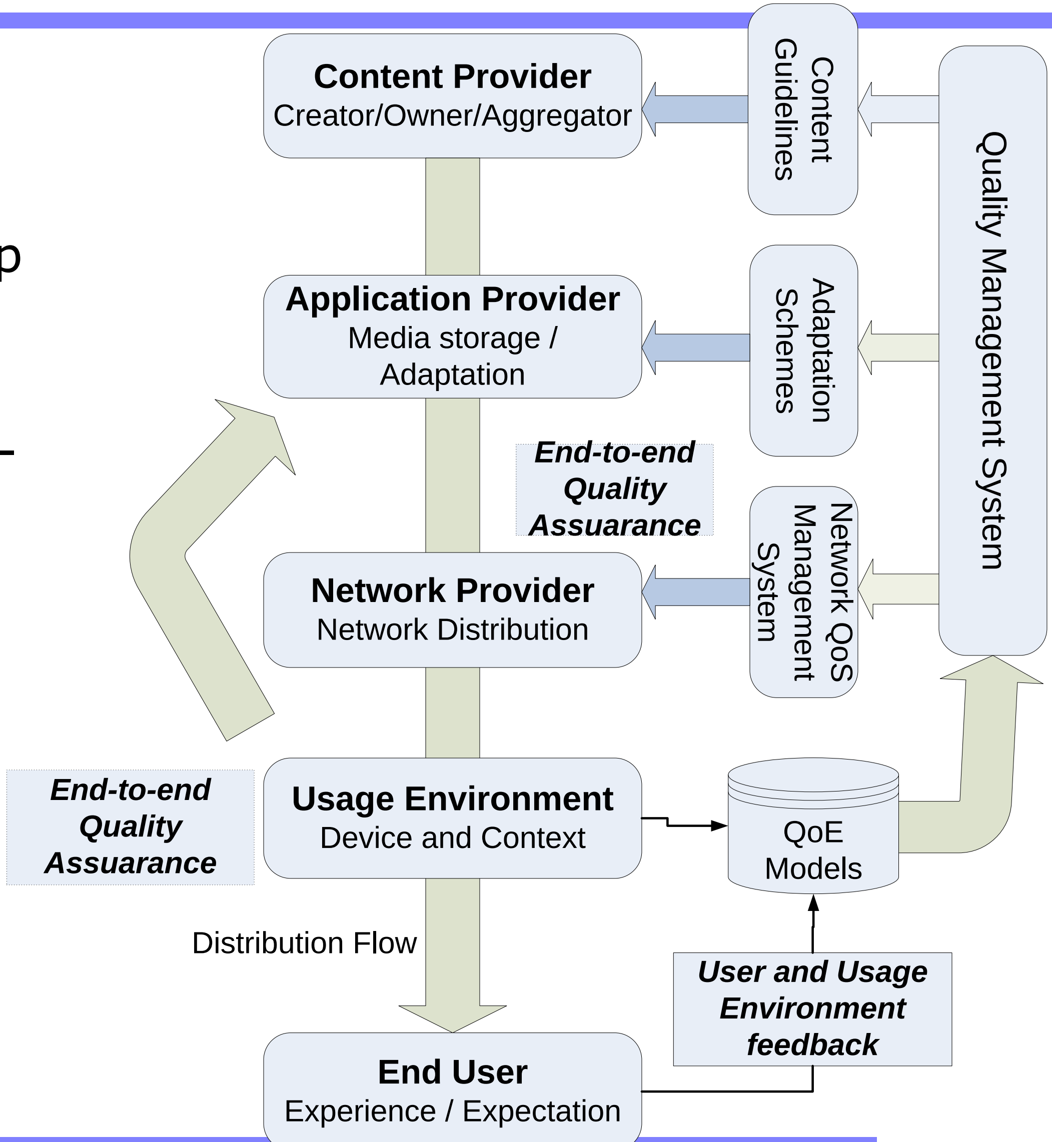
Orbi sequence



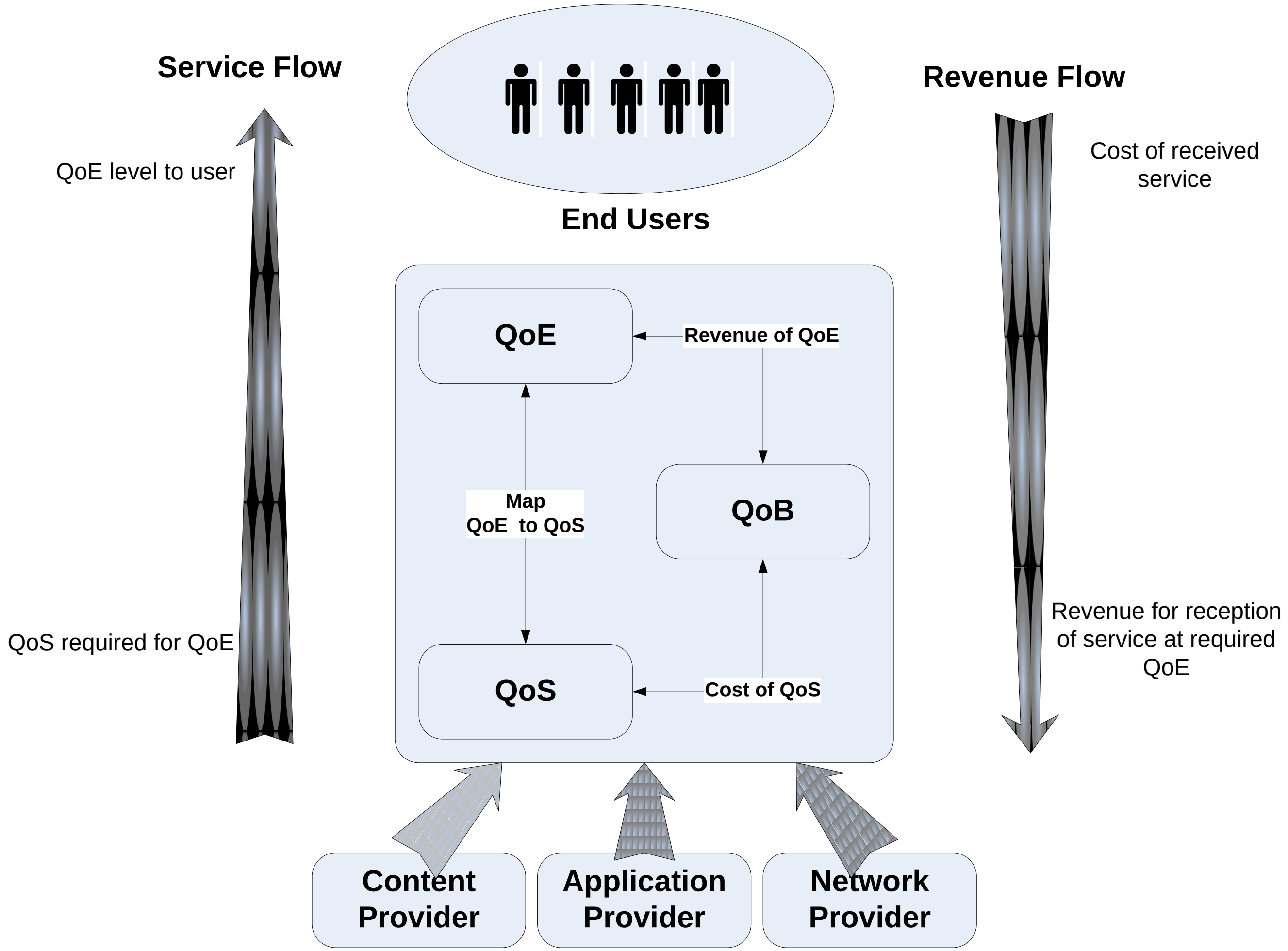
Interview sequence

Major QoE Challenges…

- The QoE concept represents the ultimate performance measure for the end-to end experience by closing the loop which includes the traditional QoS network management mechanisms
- Automated translation of end-users' QoE into objective performance measures, by identifying KPIs of different applications
- Dynamically optimised the whole delivery chain to maximise QoE

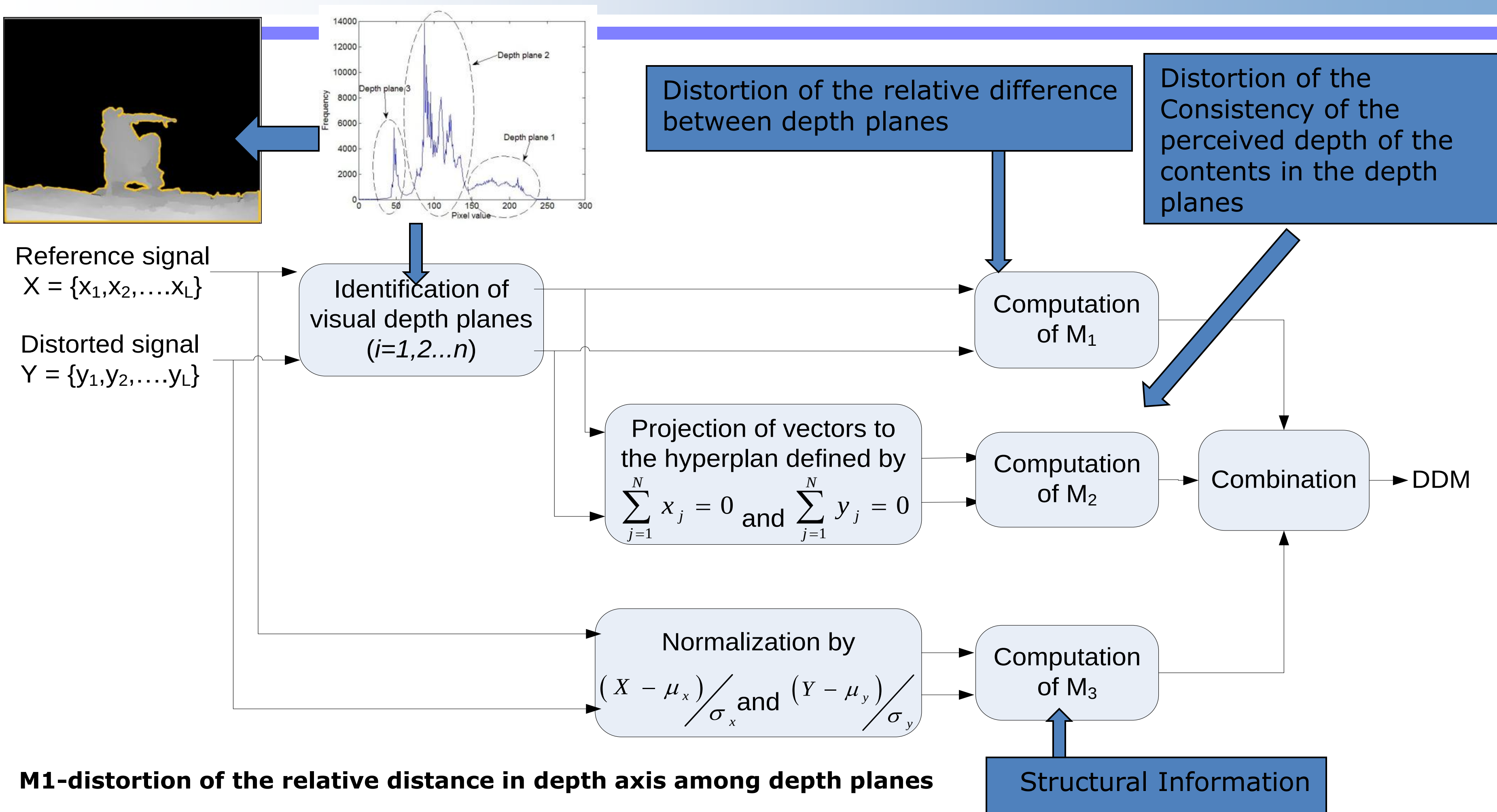


3Q concept: QoE, QoB, QoS



The interaction of QoS, QoE and QoB in 3D Networked Media Systems

Disparity Distortion Model



M1-distortion of the relative distance in depth axis among depth planes

M2- distortion of the consistency of the perceived depth of the content in the depth planes

M3- structural error of the depth map

UNIVERSITY OF SURREY

Indicative Results

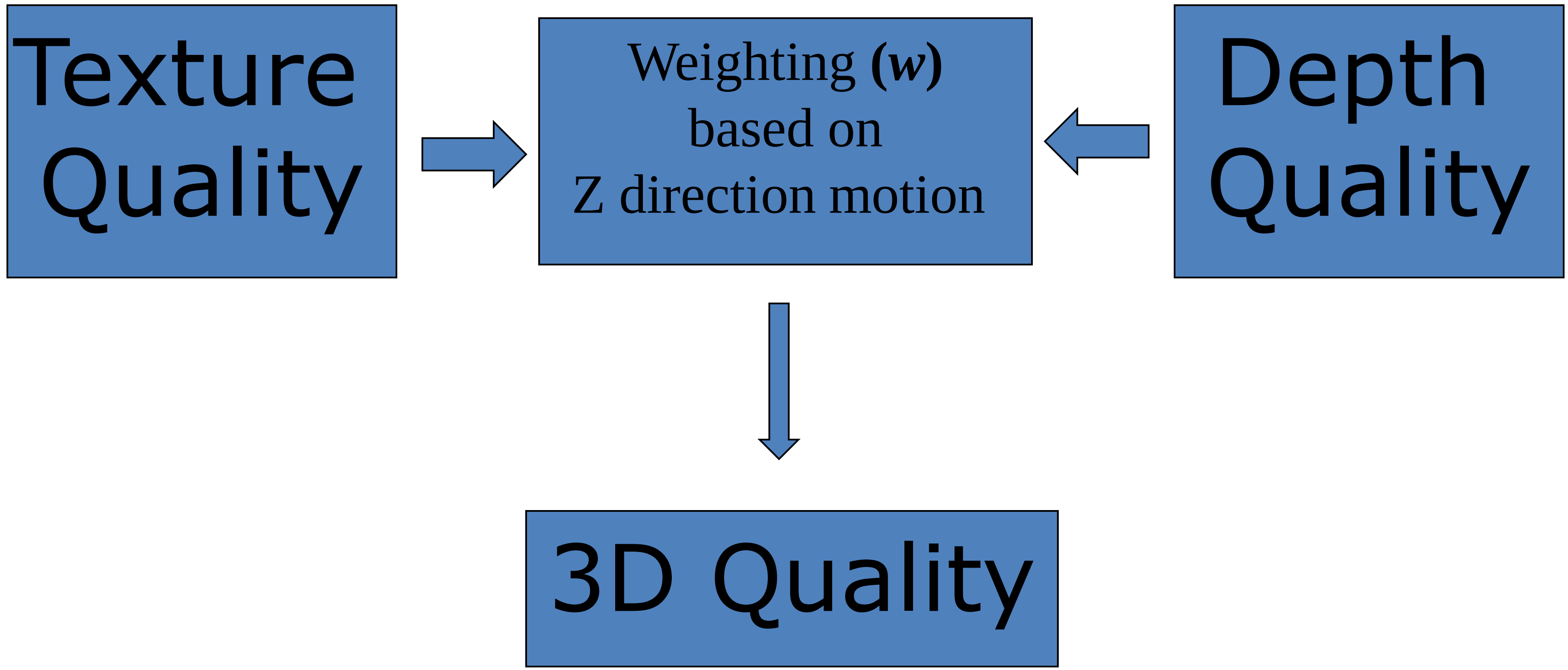
- **Performance of the proposed model: tested according to ITU-Recommendations (ITU-R BT.1438)**
  - Cc=correlation coefficient, RMSE= root mean squared error, SSE = sum of squared error

Objective Quality Model	Overall depth perception		
	CC	RMSE	SSE
Average PSNR of the Rendered Left and Right views	0.7788	0.0737	0.0579
Average SSIM of the Rendered Left and Right views	0.8065	0.0674	0.0547
Average VQM of the Rendered Left and Right views	0.7753	0.0739	0.0603
Proposed Model	**0.8708**	**0.0328**	**0.0382**

Normalized: CC=1, RMSE=0 and SSE=0 perfect correlation
CC=0, RMSE=1 and SSE=1 worst correlation

UNIVERSITY OF SURREY

Overall 3D Video Quality



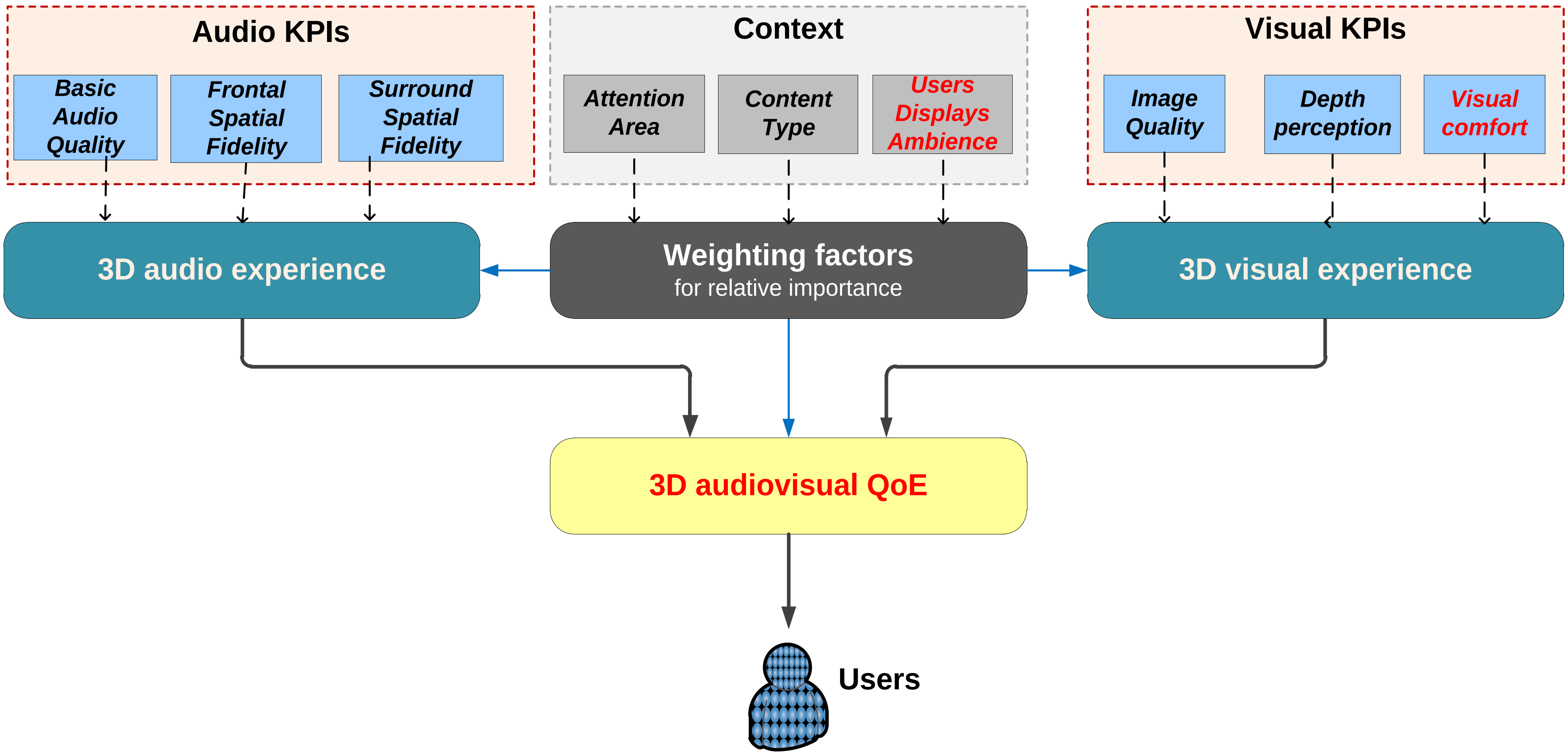
$$3D\,Quality = w.Texture\;Quality + (1-w)Depth\;Quality$$

UNIVERSITY OF SURREY

3D Video Quality Results

Objective Quality Model	Overall 3D Quality		
	CC	RMSE	SSE
Average PSNR of the Rendered Left and Right views	0.7061	0.1363	0.1091
Average SSIM of the Rendered Left and Right views	0.7387	0.0949	0.0887
Average VQM of the Rendered Left and Right views	0.8092	0.0570	0.0501
Proposed Model	0.8441	0.0347	0.0328

3D audio-visual QoE model



Final Comments…

- Content processing forms a very important part of the current networks and will gain even more importance in the future as User’s will choose their applications (push+pull)

- Inclusion of all aspects of content from capture to consumption is a must for the current and future network planners

Thank you for listening!