KINGSTON UNIVERSITY VIDEO QUALITY RESEARCH

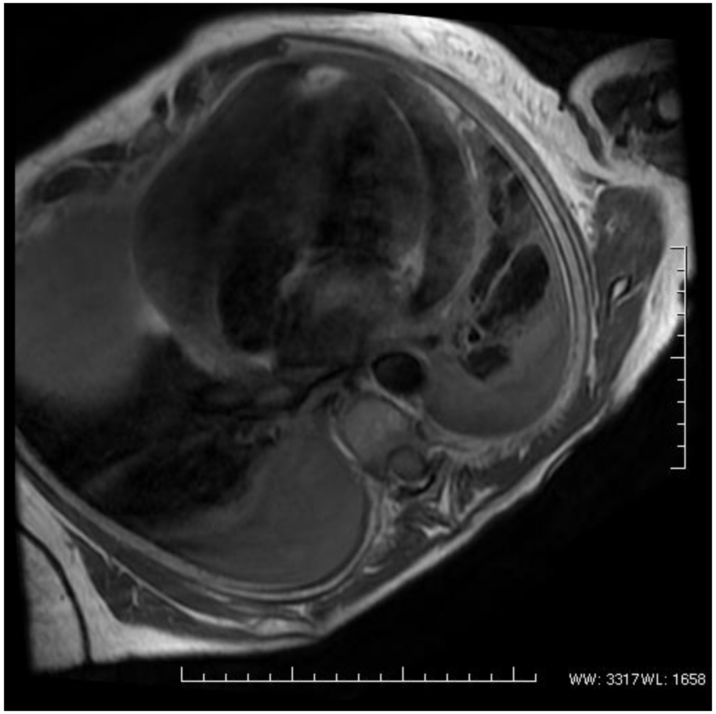
Prof Maria Martini
Kingston University, London, UK

A Study on Quality Assessment for Medical Ultrasound Video Compressed via HEVC

Manzoor Razaak, Maria Martini, Ketty Savino

Medical Image/Video Quality Evaluation

What makes it more challenging?



- In medical images **the diagnostic quality** of the image is more important than the perceptual quality
- However, no objective image quality metric measures the diagnostic quality

Medical Image quality evaluation
Objective Quality Measurement

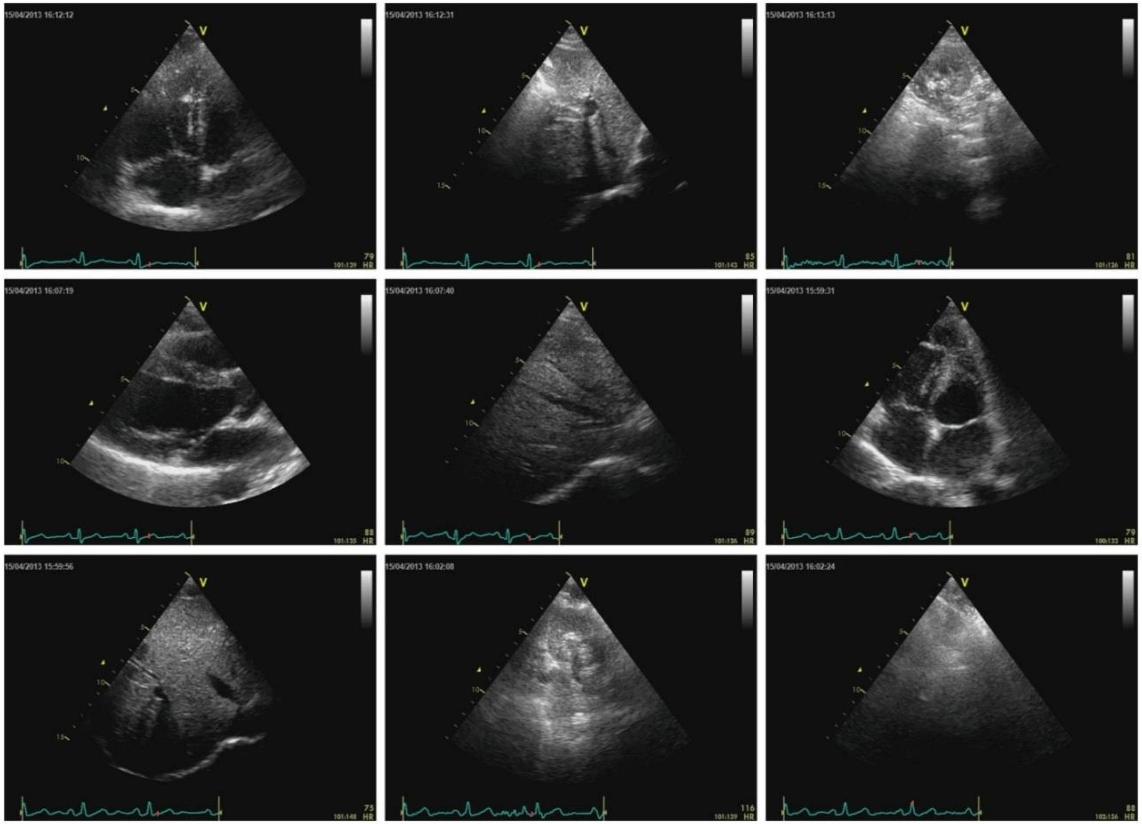
Objective Image Quality Metrics	Defining Equation		Features considered
PSNR	$PSNR = 20 * \log_{10}\left(\frac{255}{\sqrt{MSE}}\right)$	(1)	Gives ratio of signal over the noise, where the noise is represented in terms of mean square error (MSE).
SSIM [13]	$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$	(2)	Measures the structural similarity between two images. The constant values are chosen to be $C_1 = 0.01, C_2 = 0.03$.
universal quality index (UQI) [16]	$Q = \frac{4\sigma_{xy}\mu_x\mu_y}{(\sigma_x^2 + \sigma_y^2)(\mu_x^2 + \mu_y^2)}$	(3)	Measures the structural distortion and gives good approximation of perceived distortion.
VQM [17]	A standardized metric developed by the National Telecommunications and Information Administration, USA. VQM is also recommended by ITU and is widely adopted. VQM measures the quality based on seven different parameters. The software is freely available at [17].		
noise quality metric (NQM) [18]	NQM is a weighted SNR measure between the original and the processed image. NQM considers variation in contrast sensitivity, local luminance mean, and contrast measures.		
visual information fidelity (VIF) [19]	The images are decomposed into wavelets and computation is done using several models which gives a measure of the visual quality.		
VSNR [14]	Contrast thresholds are used to identify distortions. The distortions above the threshold are modeled to represent measure the image quality.		

How well do these metrics correlate with subjective results?

Extensive subjective tests run with medical doctors in Hospital of Perugia

Medical image quality evaluation

Dataset



- Nine ultrasound video sequences provided by cardiologist in Perugia Hospital (Heart, Liver, Kidney, Lung), 640 x 416, 100 frames, 25fps
- Compressed with HEVC at eight different compression ratios

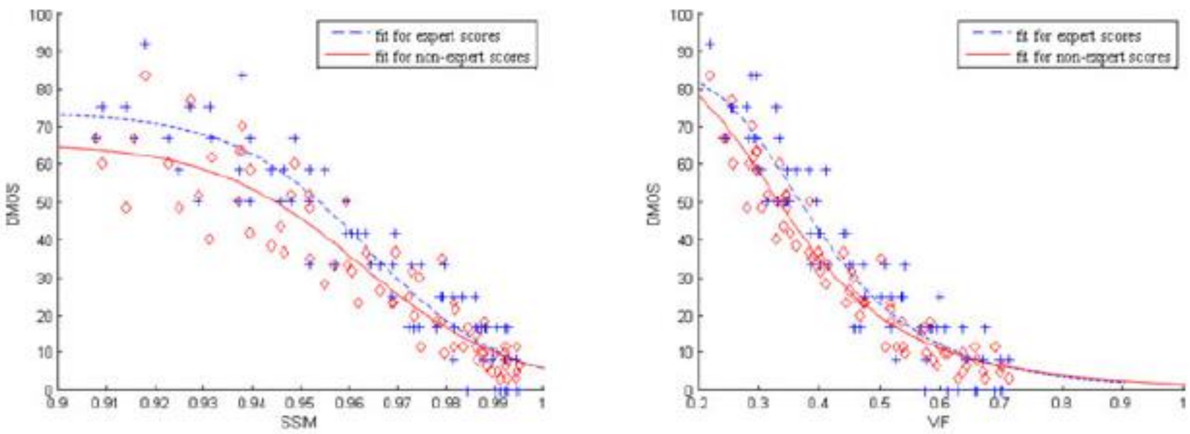
9 x 8 = 72 video sequences being evaluated with DSCQS (**144 sequences evaluated by each specialist** not including
intra-subject controls)

Medical image quality evaluation
Example results

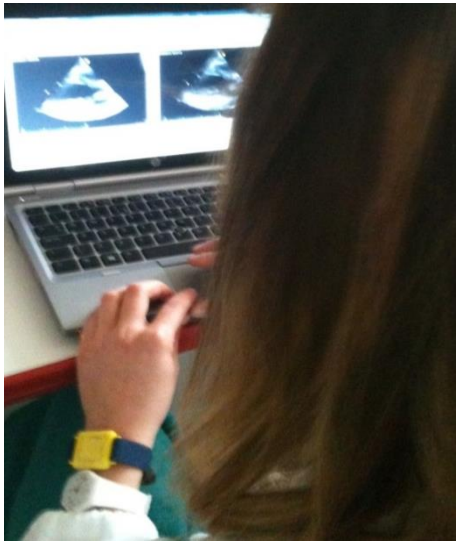
- Tests run at Hospital, where cardiologists rated the sequences based on diagnostic accuracy
- Tests on the same sequences also run with non-experts based on perceived video quality

Correlation of DMOS with different objective metrics evaluated

Example scatter plot:



(a) (b)



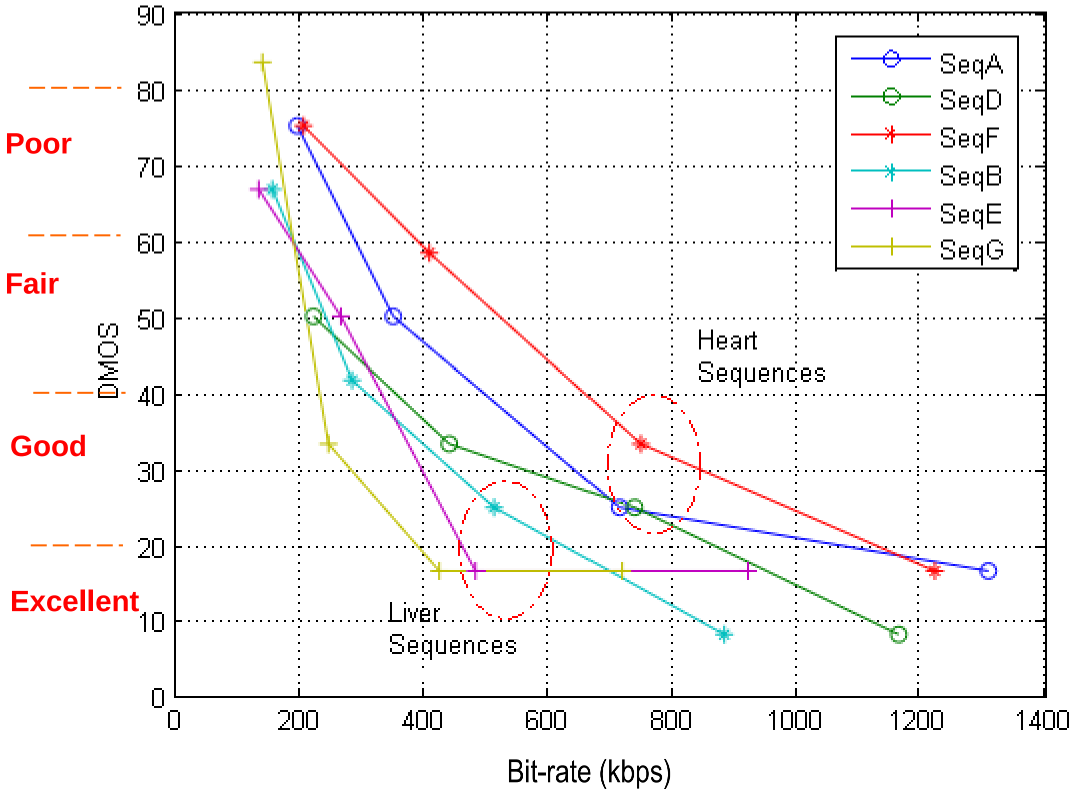
Results - 1

Accuracy comparison

	CC	State of the art objective metrics						
		SSIM	VSNR	VIF	UQI	PSNR	VQM	NQM
Experts	PLCC	0.9264	0.8925	0.9258	**0.9292**	0.9109	0.8080	0.8961
	$PLCC_{Nlin}$	0.9366	0.9150	**0.9417**	0.9309	0.9261	0.8309	0.9127
	SROCC	0.9375	0.9139	**0.9382**	0.9251	0.9331	0.8368	0.9090
Nonexperts	PLCC	0.9208	0.8888	0.9431	**0.9521**	0.8896	0.8146	0.9233
	$PLCC_{Nlin}$	0.9279	0.9200	**0.9688**	0.9521	0.9116	0.8440	0.9435
	SROCC	0.9383	0.9277	**0.9663**	0.9495	0.9280	0.8606	0.9464
Experts and Nonexperts combined	PLCC	0.9350	0.9022	0.9532	**0.9612**	0.9064	0.8249	0.9312
	$PLCC_{Nlin}$	0.9427	0.9321	**0.9770**	0.9612	0.9272	0.8532	0.9510
	SROCC	0.9570	0.9456	**0.9801**	0.9633	0.9475	0.8789	0.9587

Results - 2

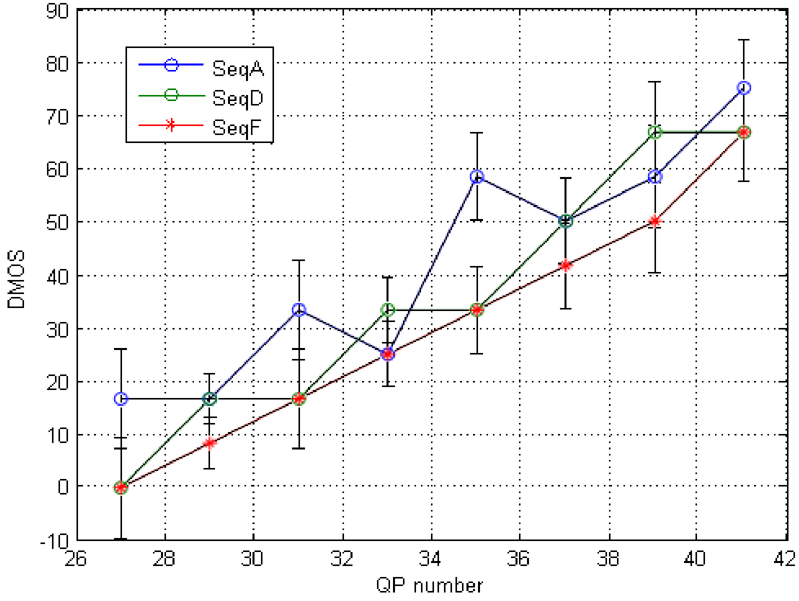
Rate-Quality Analysis



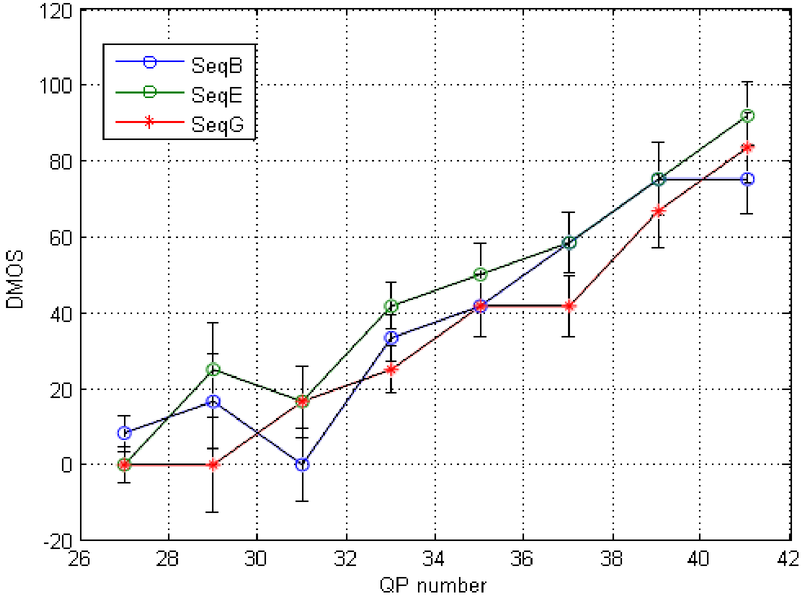
- “Excellent” diagnostic quality video is obtained at different bitrates for different sequence content
- The compression ratios for “excellent” quality videos is in the range of 140:1 to 420:1
- The spatio-temporal complexity influences the compression ratio

Results - 3

Compression vs. Perceived quality



Heart sequences



Liver sequences

An acceptable diagnostic quality for medical ultrasound videos can be obtained with a maximum QP range (HEVC) of 32 - 35

Observations and further work

- The specialists clearly accepted compression and were often not able to distinguish compressed and original video sequence
- Quality rating on diagnostic value depends on experience
- Edges (and motion for cardiac ultrasound) are critical for diagnosis

- More subjective results would enable higher confidence in results
- Further statistical analysis of data
- Intra-subject consistency can be further investigated
- Different test methodologies can be adopted

References

M. Razaak, M.G. Martini and K. Savino, "A Study on Quality Assessment for Medical Ultrasound Video Compressed via HEVC," *IEEE Journal of Biomedical and Health Informatics (J-BHI)*, vol. 18, no. 5, pp. 1552-1559, Sep 2014.

M. Razaak and M.G. Martini, "CUQI: Cardiac Ultrasound Quality Index," *SPIE Journal of Medical Imaging*, 2016.

Kingston University London

3D Video Quality Database

Existing 3D image/video databases

- Mobile 3D video database by Tampere Univ. and Nokia (Crosstalk, blocking, color mismatch and bleeding, packet losses, etc.)
  - No associated MOS values
- IRCCyN image database by Nantes University (JPEG, J2K, upsample/downsample, etc.)
  - No transmission impairments
- EPFL database by EPFL (different camera distances)
  - No impairments, just different camera distances
- RMIT3DV Database (uncompressed HD3D video)
- LIVE database (only 3D images, distortion types JP2K, JPEG, WN, FF, Blur)
  - No packet losses
- Current IEEE P3333.1.2 database

Objectives

- Design a stereoscopic video database for the evaluation of 3D visual quality assessment metrics for images and video affected by packet losses
- Evaluate the effects of random packet losses on the overall 3D perception

↓

- We developed a database of 3D video sequences affected by H.264 compression and packet losses with different ratios and distribution
- Each impaired sequence is included in the database together with the **MOS** obtained in subjective tests

Impairments

- H.264 compression at different compression levels

1) H.264 compression + Packet losses (equiprobable for each packet - binary erasure channel)
2) H.264 compression + Packet losses according to Gilbert-Elliot model / LTE loss traces

3D Video dataset 1

- Nine left and right views based 3D video sequences (i.e., *café, beer garden, ballroom, newspaper, lovebird, kendo, horse, car and mobile*) from different sources (freely available for scientific research)
- The sequences cover a wide range of image textures, motion profiles, depth levels and frame rates (e.g., 15 fps and 30 fps) and image resolutions








Description of the Proposed 3D Video Database

- Five test sequences obtained from each reference / original 3D sequence using simulated average packet loss rates of 0%, 1%, 3%, 5% and 10%
- 3D video compression
  - H.264/AVC (*JM* Reference software version 12.0)
  - QP = 30 (to obtain good quality videos)
  - *IPPP...IPPP...* sequence
  - *I* frame is encoded by every one *second*
  - *RTP* bit-stream

Description of the Proposed 3D Video Database

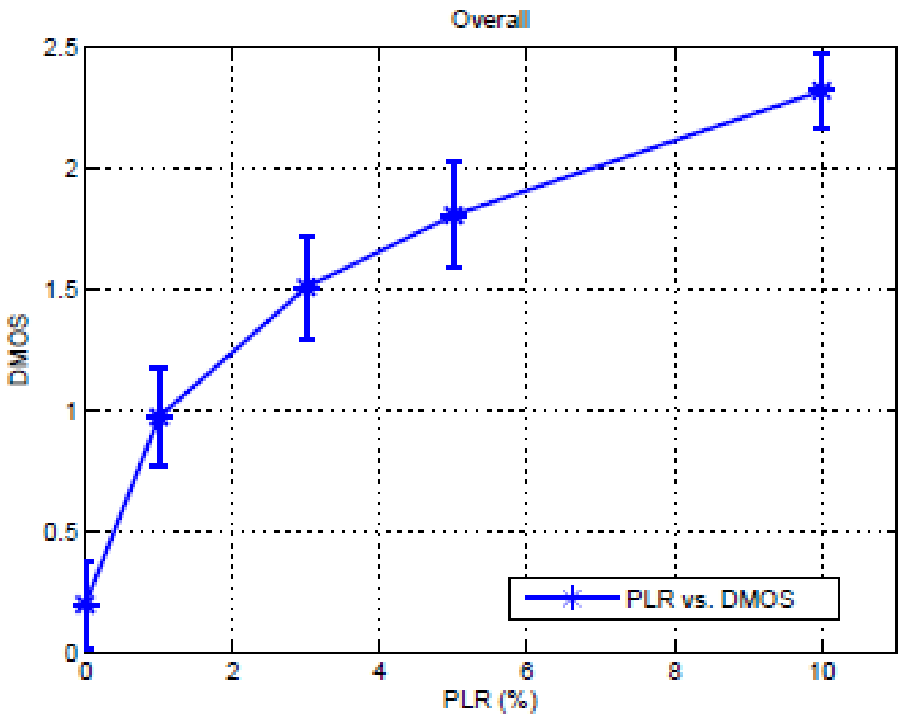
- Packet loss simulation
  - JM Reference software is employed to drop packets randomly at 0%, 1%, 3%, 5% and 10% loss rates with the same probability for each packet (memoryless erasure channel)
  - Error concealment: *slice-copy*
  - 100 simulation runs to obtain average video quality
  - Sequence close to average PSNR is selected for subjective quality evaluation results

Description of the Proposed 3D Video Database

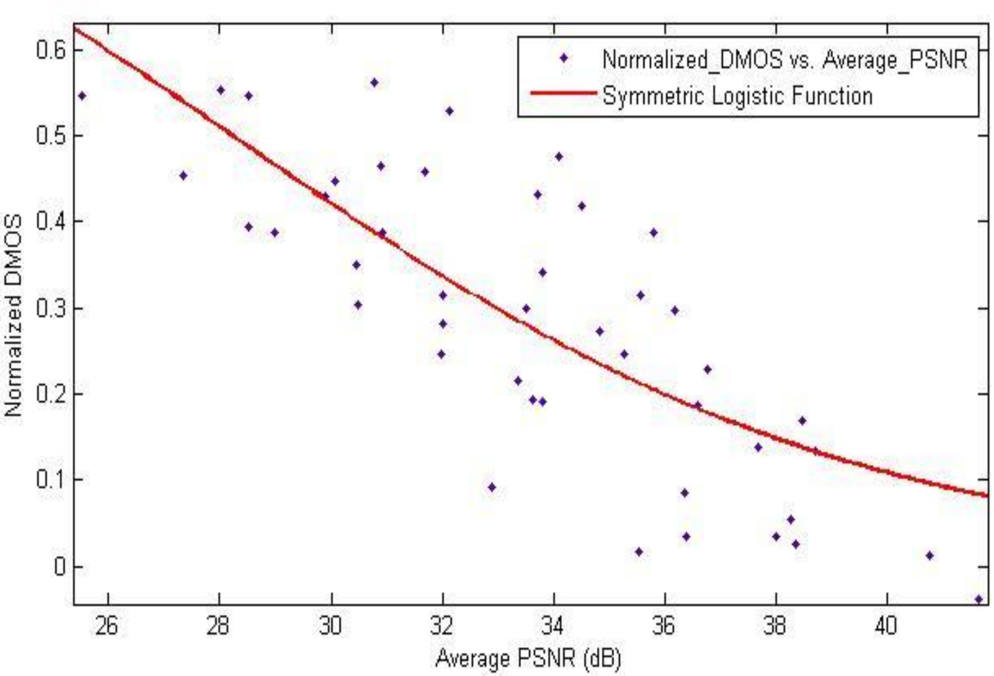
- Subjective quality evaluation experiments
  - Observers: 16 non-expert observers
  - Stimuli: 9 left and right stereoscopic video sequences, are used in this experiment
    - stimulus set of 9 x 6 = 54 3D video sequences
  - Equipment: The 47'' LG Polarized stereoscopic display
  - Procedure: A set of 108 stereoscopic test video sequences is randomized and presented sequentially
  - DSCQS tests

Subjective quality ratings

DMOS VS. PLR, FOR ALL THE SEQUENCES



OBJECTIVE QUALITY



DB2 - Snapshots of the 3D video test sequences

(a) *barrier.*

(b) *news-readers.*

(c) *football.*

(d) *umbrella.*

DB2 – LTE Simulation setup

Parameters	Values
Simulation time	30 s
Video flow duration	20 s
Frame structure	FDD
Cell radius	1 km
E-UTRAN frequency band	1 (2.1 GHz)
Bandwidth	10 MHz
Scheduler type	M-LWDF
Max delay	100 ms
3D video bit-rate	2.2 Mbps
Propagation loss/ Channel model	Typical Urban /Pedestrian A
Simulation repetitions	10

Quality Assessment

- Objective metrics
  - PSNR
  - SSIM
- Subjective tests
  - Tests run in WMN Research Group LAB at Kingston University
  - 15 naive subjects
  - LG 47" HD-3D display with polarized glasses
  - Absolute Category Rating with Hidden Reference (ACRHR) method
  - The impaired sequences for subjective quality testing are chosen to represent the average quality of several simulation runs
  - Results analysed according to ITU-R BT.500-13 Recommendation
  - Mean opinion scores and 95% confidence intervals (CIs) are calculated.
  - No outliers were found for the tests considered.

Results

- Objective quality results:
  - average PSNR and SSIM quality results.

3D sequence	Average PSNR (dB)						
	Original	*One user*	*Two users*	*Three users*	*Four users*	*Five users*	*Six users*
barrier	34.39	32.74	32.42	30.53	25.20	25.10	22.05
news-readers	38.00	36.82	35.08	31.04	26.19	24.90	22.98
football	35.02	32.46	30.73	26.74	24.08	23.00	20.00
umbrella	31.15	29.48	28.41	26.76	24.34	23.37	20.00
3D sequence	**Average SSIM**						
barrier	0.89	0.88	0.88	0.86	0.83	0.82	0.78
news-readers	0.95	0.94	0.94	0.93	0.90	0.88	0.87
football	0.93	0.92	0.90	0.86	0.78	0.79	0.66
umbrella	0.84	0.83	0.82	0.81	0.77	0.76	0.61

Results

- Subjective quality results:
  - the table shows the MOS, the standard deviation and the 95% confidence interval for the selected 3D test video sequences (as recommended in Annex 2 of ITU-R BT.500-13).

Sequence	*umbrella*							*barrier*						
Number of users	*Original*	*One user*	*Two users*	*Three users*	*Four users*	*Five users*	*Six users*	*Original*	*One user*	*Two users*	*Three users*	*Four users*	*Five users*	*Six users*
Mean	4.26	2.8	2.53	2.46	2.06	1.33	1.06	4.93	3.6	3.53	2.73	2.53	2.46	1.33
STD	0.59	0.77	0.51	0.83	0.70	0.61	0.25	0.25	0.50	0.51	0.88	0.83	0.51	0.48
95% CI	0.30	0.39	0.26	0.42	0.35	0.31	0.13	0.13	0.25	0.26	0.44	0.42	0.26	0.24
Sequence	*football*							*news-readers*						
Number of users	*Original*	*One user*	*Two users*	*Three users*	*Four users*	*Five users*	*Six users*	*Original*	*One user*	*Two users*	*Three users*	*Four users*	*Five users*	*Six users*
Mean	4.53	3.33	3.26	2.2	2	1.86	1.06	4.8	3.53	3.4	2.26	1.53	1.4	1.06
STD	0.51	0.97	0.70	0.67	0.75	0.63	0.25	0.41	0.51	0.63	0.70	0.63	0.63	0.25
95% CI	0.26	0.49	0.35	0.34	0.38	0.32	0.13	0.20	0.26	0.32	0.35	0.32	0.32	0.13

References

- C.T. Hewage, M.G. Martini, M. Brandas, and D. V. S. X. De Silva, “A study on the perceived quality of 3D video subject to packet losses” Proc. *IEEE International Conference on Communications,* Budapest, June 2013 (pp. 662-666).
- C.T. Hewage, M.G. Martini, "Quality of Experience for 3D video streaming“, *IEEE Communications Magazine*, vol. 51, no. 5, pp. 101-107, May 2013.
- M. M. Nasralla, C. T. Hewage, and M. G. Martini, “Subjective and objective evaluation and packet loss modeling for 3D video transmission over LTE networks”, *IEEE International Conference on Telecommunications and Multimedia (TEMU), July 2014* (pp. 254-259).

3D and 2D RR video quality metrics

C.T. Hewage and M.G. Martini, "Quality of Experience for 3D video streaming," *IEEE Communications Magazine*, May 2013.

C.T. Hewage and M.G. Martini, "Edge based Reduced-Reference Quality Metric for 3D Video Compression and Transmission," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 471-482, 2012.

C.T. Hewage, M.G. Martini, *"*Reduced-Reference Quality Assessment for 3D Video Compression and Transmission*"* in *IEEE Transactions on Consumer Electronics*, August 2011.

M.G. Martini, C.T. Hewage, B. Villarini, *"*Image Quality Assessment based on Edge Preservation*"* in *'Signal Processing: Image Communication'*, 27(8), Elsevier, September 2012, pp. 875–882.

QoE for MPEG-DASH

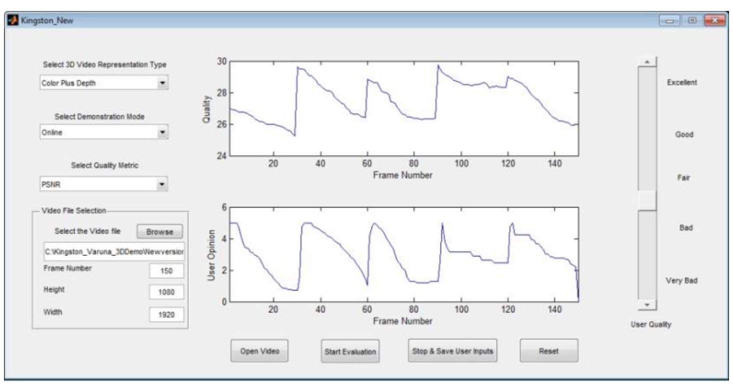
Continuous quality monitoring

Kingston University continuous quality monitoring tool

- Records quality variations over time
- Can support video streaming (live/VoD) and local playback

2 versions:

- Matlab
- Android application



M.G. Martini, C.T. Hewage et al. "Real-time multimedia communications in medical emergency-the CONCERTO project solution" , *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015.

Perceived effect of frame freezing

How does the frame freezing **tolerance** vary over time for frames with different Temporal Information (TI) values?

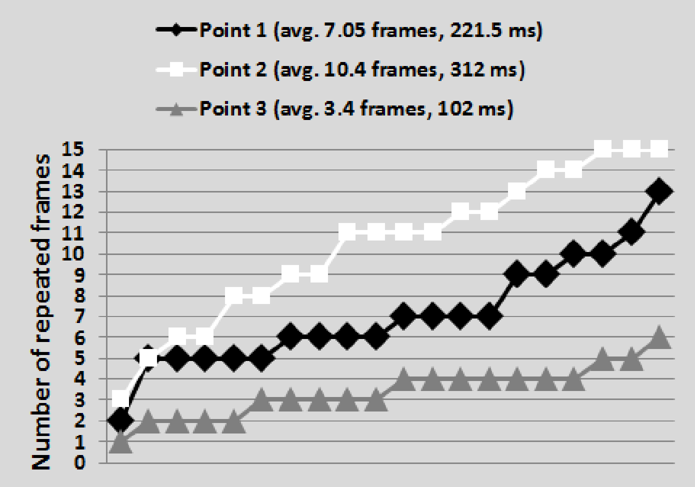
What is the significance of movements on the 2D plane and in the third dimension (back–forth movement) during frame freezing events?

Investigation of the **perceptibility** of frame freezing, assessing whether test participants can detect video freezing or not, regardless of their level of perceptual tolerance.

Results not only from the angle of mean scores, but also from scoring distributions of individual participants.

P. A. Kara, W. Robitza, M.G. Martini, C.T. Hewage and F. M. Felisberti, "Getting Used to or Growing Annoyed: How Perception Thresholds and Acceptance of Frame Freezing Vary Over Time in 3D Video Streaming" in *IEEE International Conference on Multimedia and Expo (ICME) 22nd International Packet Video Workshop (PV)*, Seattle, USA, Jul 11-15 2016.

P. A. Kara, M.G. Martini and S. Rossi, "One Spoonful or Multiple Drops: Investigation of Stalling Distribution and Temporal Information for Quality of Experience over Time" in *IEEE International Conference on Telecommunications and Multimedia (TEMU)*, Heraklion, Crete, Greece, Jul 25-27 2016, pp. 157-162



Content characterization for encoding in appropriate representations

- Adaptive video streaming adapts the bitrate by selecting a representation with different video resolution and compression parameters depending on the user's available bandwidth
- The required encoding bitrate and resolution varies with the video complexity and to a certain extent with the metric used for quality evaluation.

- Content-based video classification solution to classify the videos in different content classes

Experimental Biases in User Behavior and QoE Assessment in the Lab

- Observation and measurement of user behavior during streaming multimedia consumption in a controlled laboratory environment.
- Subjects were asked to watch a number of video sequences and describe their contents.
- During the process, they were confronted with typical video streaming problems such as stalling and quality fluctuations, which often result in a low Quality of Experience for end-users.
- The subjects were not informed about the real purpose of the test and their behavior was tracked unobtrusively.
- The results suggest that the method can indeed elicit realistic responses such as seeking, pausing, or reloading the web page. A third of subjects acted apprehensively, meaning that they changed their behavior due to being part of a test.

W. Robitza, P. A. Kara, M. G. Martini, A. Raake, “Experimental Biases in User Behavior and QoE Assessment in the Lab”, IEEE Globecom 2016 – QoEMCWorkshop, Dec 2016.

Content-aware video quality model for MPEG-DASH

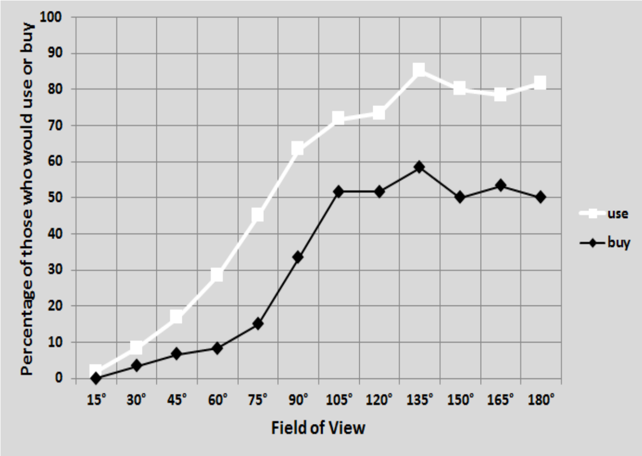
- First draft finalised
- Under test

M.G. Martini| October 2016

QoE for Light Field Displays

Subjective assessment of field of view for light field displays

- Completed by 20 test participants
- Subjective evaluation of the different field of views, ranging from 15 to 180 degrees
- Decisions regarding display usage and purchase



P. A. Kara, P. T. Kovacs, M.G. Martini, A. Barsi, K. Lackner and T. Balogh, "From a Different Point of View: How the Field of View of Light Field Displays affects the Willingness to Pay and to Use" in *8th International Conference on Quality of Multimedia Experience (QoMEX)*, Lisbon, Portugal, Jun 6-8 2016.

P. A. Kara, P. T. Kovacs, M.G. Martini, A. Barsi, K. Lackner and T. Balogh, "Viva la Resolution: The Perceivable Differences between Image Resolutions for Light Field Displays" in *5th ISCA/DEGA Workshop on Perceptual Quality of Systems (PQS 2016)*, Berlin, Germany, Aug 29-31 2016, pp. 107-111.

HDR research

HDR

(ongoing)

Cognitive bias for HDR video

Pricing for HDR video

Tests on SIM2 display HDR47ES6MB - 6000 nits peak brightness

Acknowledgements

We acknowledge the support of European Union's

- Seventh Framework Programme (FP7) under grant agreement No. 288502 (CONCERTO) and 3DContourNet
- Horizon2020 Programme under grant agreement No. 643072 (QoE-NET)

Concerto

QoENet MSCA-ITN Network