Low-Complexity No-Reference PSNR Estimation for H.264/AVC Encoded Video

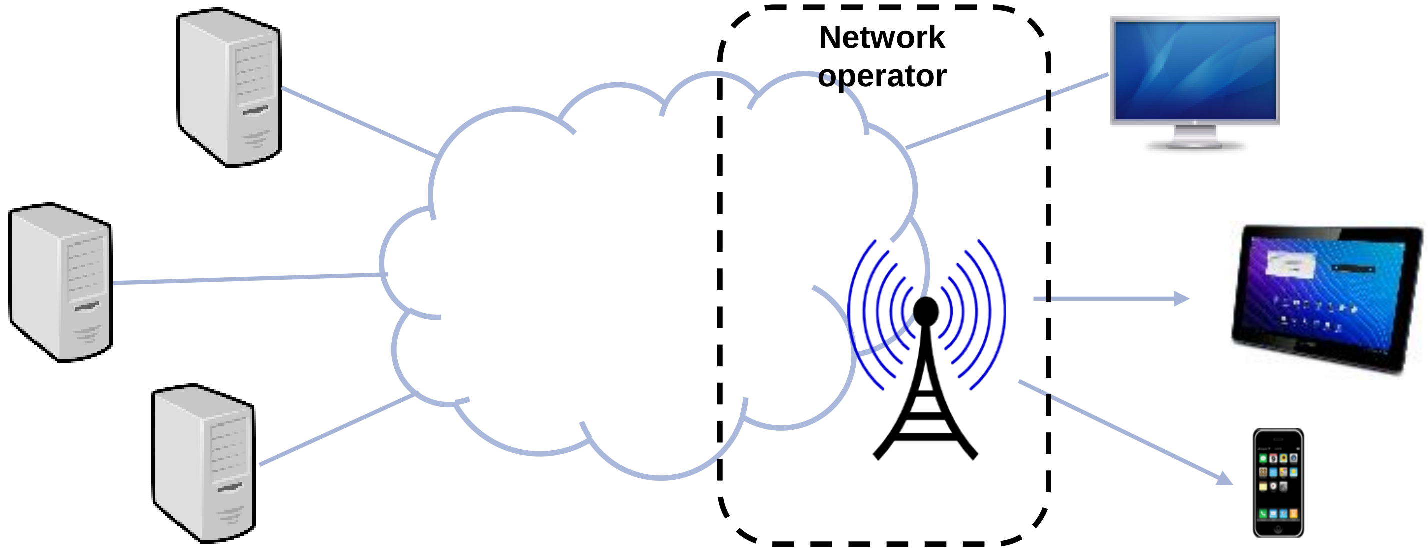
D. Schroeder[1], A. El Essaili[1], E. Steinbach[1], D. Staehle[2], M. Shehada[2]

[1]Institute for Media Technology, TUM, Munich, Germany
[2]DOCOMO Communications Laboratories Europe GmbH, Munich, Germany

Agenda

- Motivation
- Related work
- Proposed No-Reference PSNR estimation
- Validation
- Example application
- Conclusion

Motivation

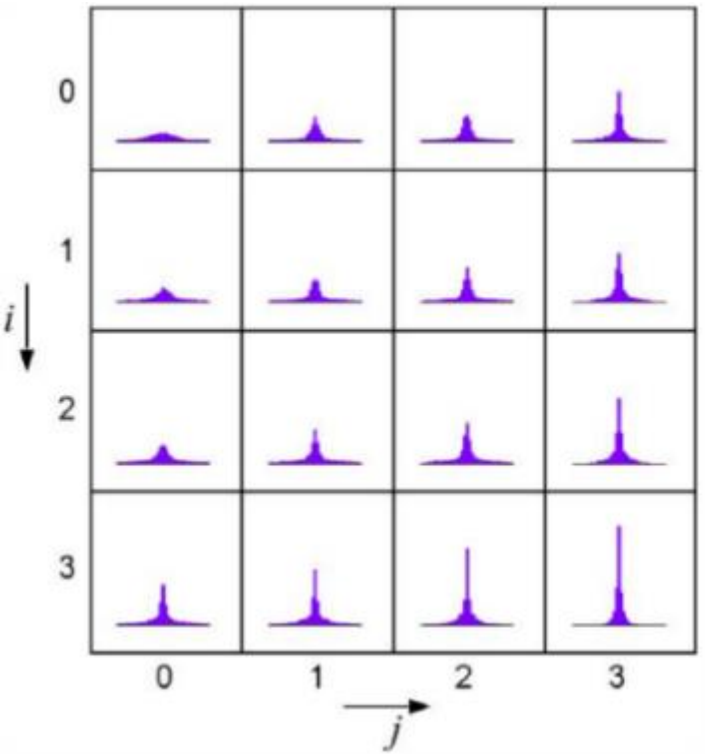


- Fast increasing video traffic to a huge number of clients
  - Need for scalable in-network quality monitoring
- Mostly HTTP/TCP based video streaming
  - No network-induced video distortion

➢ Low-complexity No-Reference PSNR estimation

Related work (1)

Statistical properties of transform coefficients



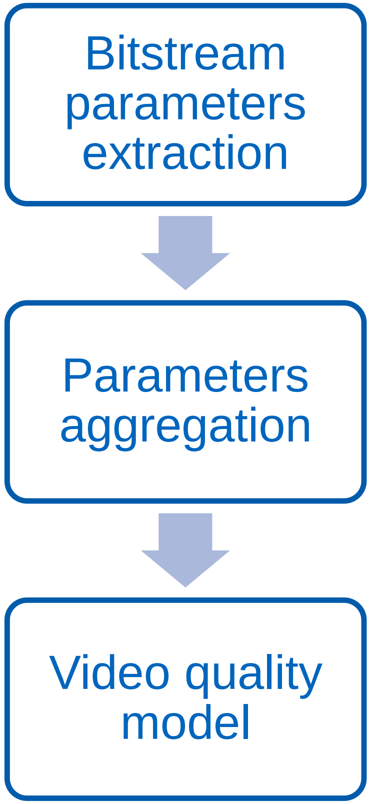
- Distribution $f_X(x)$ estimated using the quantized coefficients X_k.
- Local mean square error is estimated using: $\varepsilon_k^2 = \int_{-\infty}^{+\infty} f_X(x|X_k)(X_k - x)^2 dx$
- Estimated MSE: $MSE_{est} = \frac{1}{N}\sum_{k=1}^{N} \varepsilon_k^2$

➢ High complexity

[1] T. Brandao, M. Queluz, „No-Reference Quality Assessment of H.264/AVC Encoded Video, IEEE Trans. CSVT, 2010

Related work (2)

Bitstream features based video quality



- 21 basic parameters (motion vectors, DCT coefficients, …) are extracted from the bitstream
- Aggregated to 44 picture level parameters
- Different modules form a video quality model

- High complexity
- Targeted at RTP/UDP streaming (takes into account packet losses)

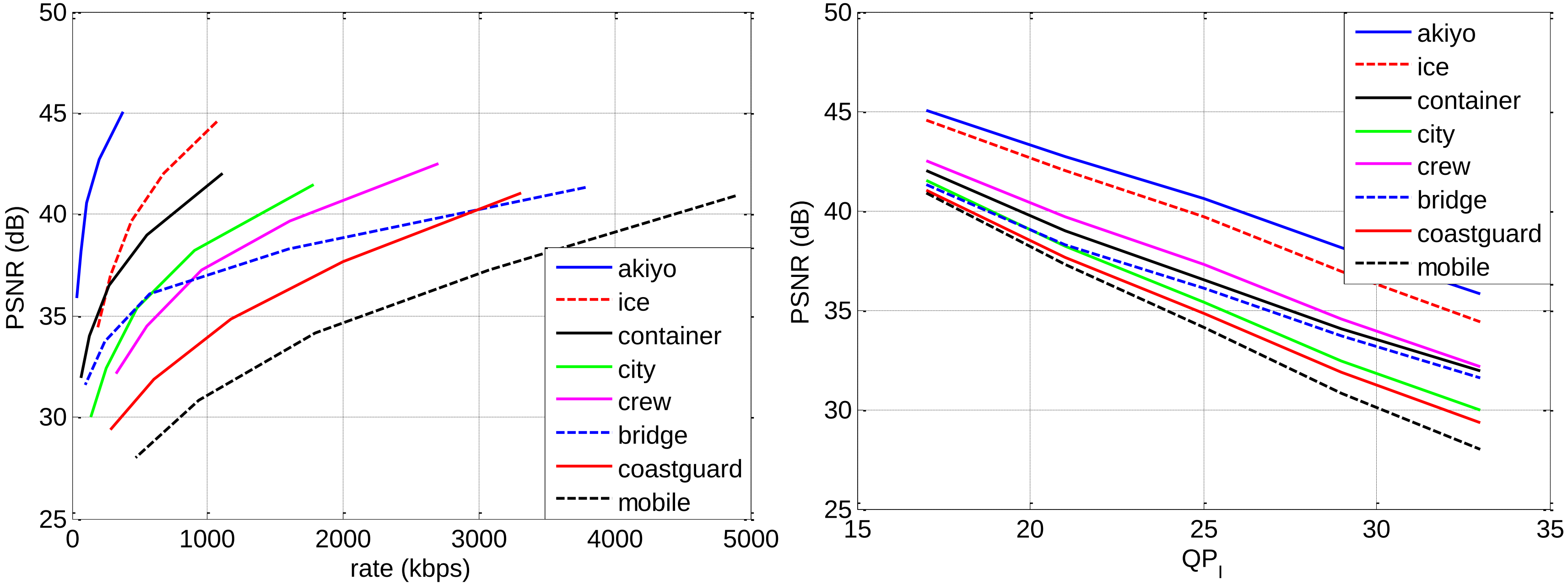
[2] ITU-T, Rec. P1202.1 „Parametric non-intrusive bitstream assessment of video media streaming quality“, 2012

Agenda

- Motivation
- Related work
- Proposed No-Reference PSNR estimation
- Validation
- Example application
- Conclusion

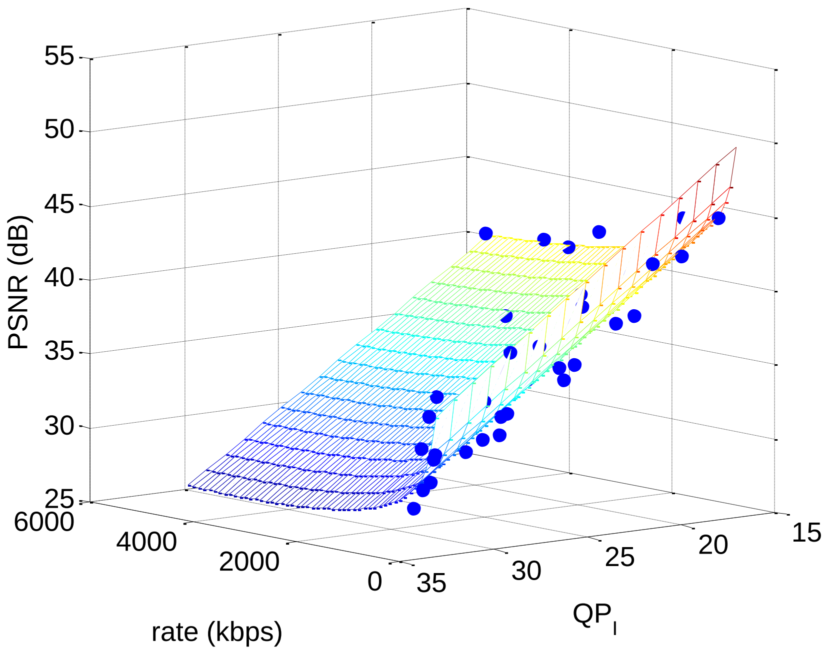
Proposed No-Reference PSNR estimation

- H.264 baseline, fixed QP
- QP_I denotes mean QP over I-frames



$$PSNR_{EST} = b_1 + b_2 * \log(rate) + b_3 * QP_I + b_4 * rate * QP_I$$

Training

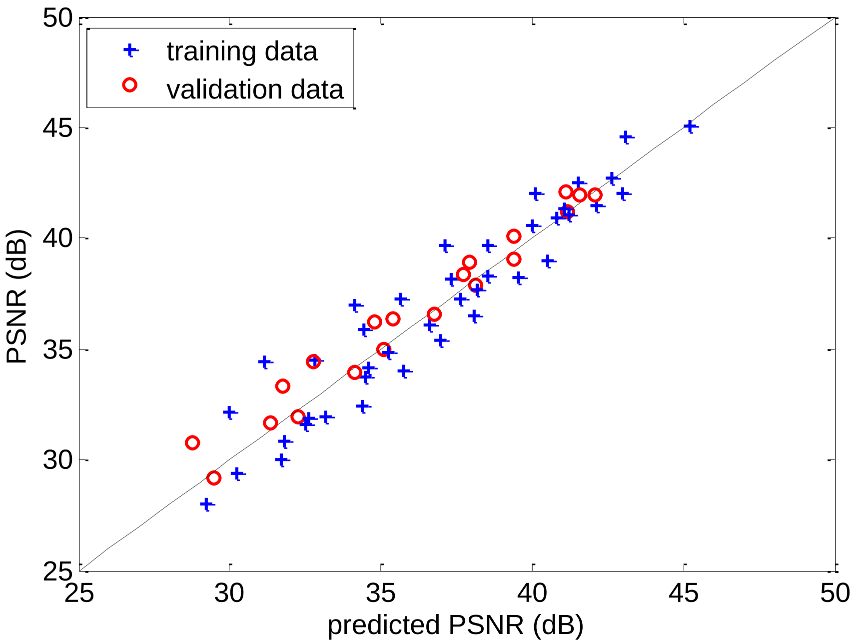


- 8 CIF videos, 30 fps
- Baseline profile
- x264 with constant QP [17, 21, 25, 29, 33]
- Multiple linear regression

$$PSNR_{EST} = b_1 + b_2*\log(rate) + b_3*QP_I + b_4*rate*QP_I$$

Model parameters	b_1	b_2	b_3	b_4
Value	74.791	-2.215	-0.975	$1.71 * 10^{-5}$

Validation (1)

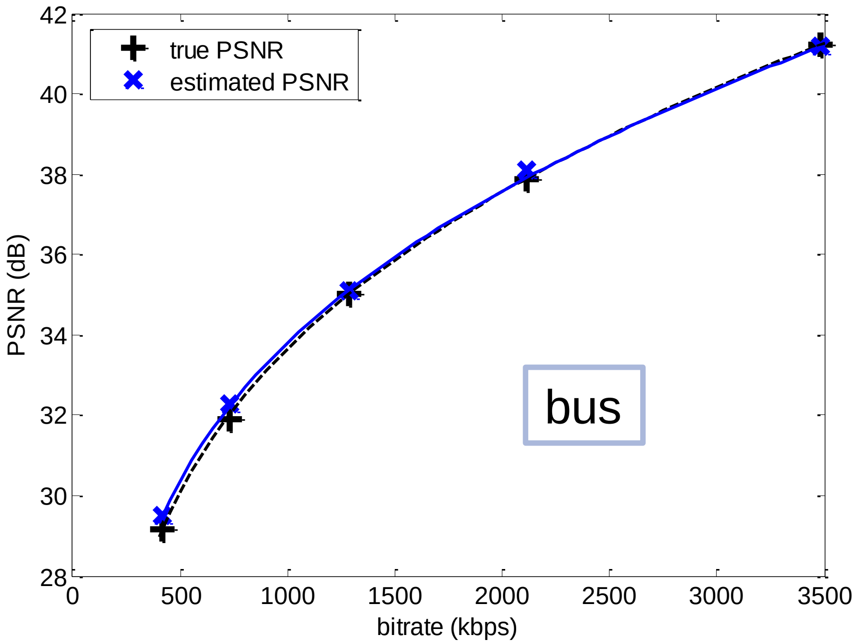


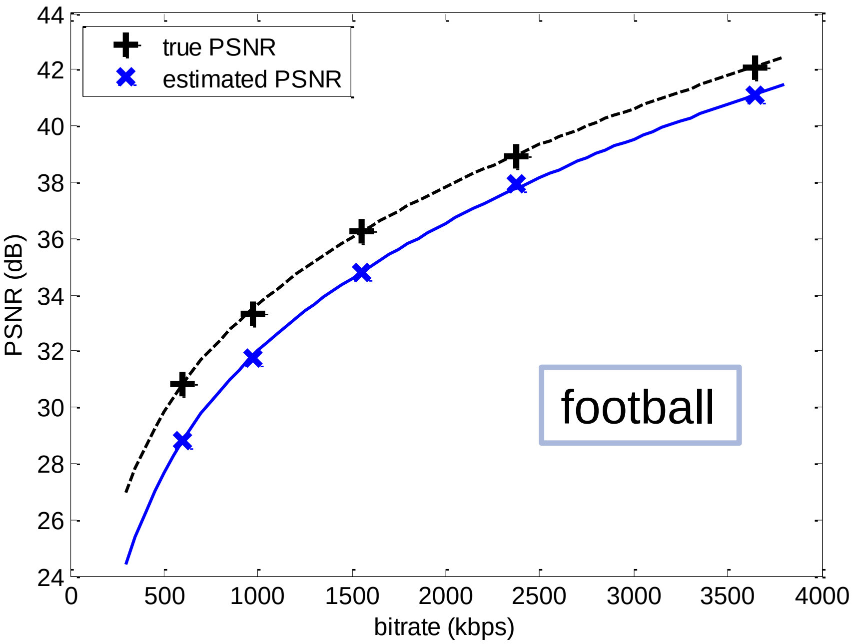
- 4 test videos: *bus, football, foreman, highway*
- Same encoding parameters as training videos

	4 validation videos	all 12 videos
RMSE [dB]	0.891	1.043
Pearson correlation	0.983	0.968

Validation (2)

Individual videos





	bus	football	foreman	highway
RMSE [dB]	0.246	1.447	0.244	0.981
PC	0.999	0.999	0.999	0.998

Influence of the encoder configuration

3-way ANOVA on the PSNR

Source	Mean Sq.	df	F-value	p-value
QP_I	378.4	1	321.51	<0.0001
bitrate	8.06	1	6.85	0.0101
mode	0.69	2	0.59	0.5569
QP_I : bitrate	18.43	1	15.65	0.0001
Error	1.177			

- Constant QP [17 … 33]
- Rate control [500 kbps … 2.5 Mbps]
- crf [19 … 27]

- 8*3*5 = 120 videos

	rate control	**crf**
RMSE [dB]	1.078	1.147
Pearson correlation	0.979	0.950

Influence of the encoding profile

3-way ANOVA on the PSNR

Source	Mean Sq.	df	F-value	p-value
QP_I	1240.19	1	542.2	<0.0001
bitrate	61.87	1	26.65	<0.0001
profile	0.97	2	0.49	0.8116
QP_I : bitrate	118.04	1	50.85	<0.0001
Error	2.32			

- x264 constant QP
  - Baseline
  - Main
  - High
- 12*3*5 = 180 videos

	main	high
RMSE [dB]	1.163	1.168
Pearson correlation	0.964	0.963

Influence of the encoder
3-way ANOVA on the PSNR

Source	Mean Sq.	df	F-value	p-value
QP_I	910.16	1	390.62	<0.0001
bitrate	43.33	1	18.6	<0.0001
encoder	3.18	1	1.36	0.2455
QP_I : bitrate	74.44	1	31.95	<0.0001
Error	2.33			

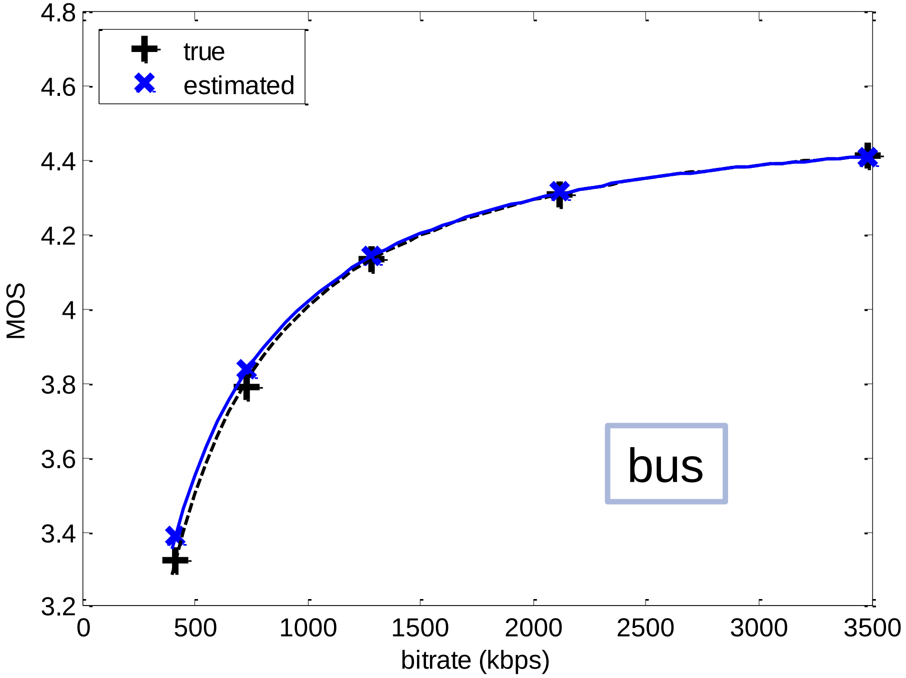
- Baseline constant QP
- JM encoder
- 12*2*5 = 120 videos

	JM encoder
RMSE [dB]	1.203
Pearson correlation	0.972

Example application (1)

- Estimation of utility curves
- Applicable to DASH videos, with multiple representations
- PSNR to MOS using [3]

➢ High PC for individual videos leads to good shape estimation



[3] S. Wolf and M. Pinson, „Video quality measurement techniques", NTIA, TR-02-392, 2002

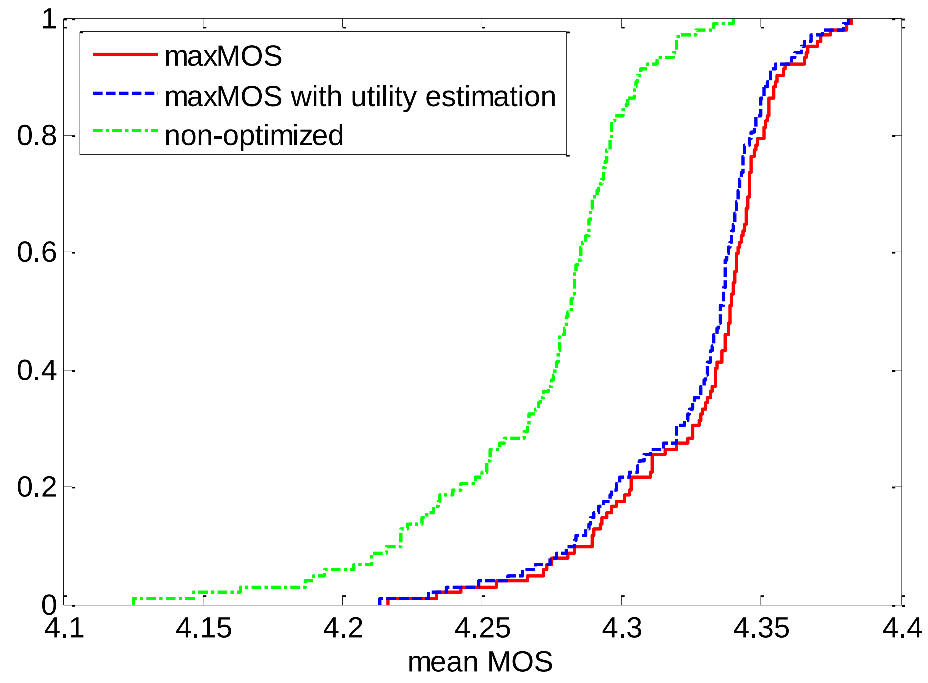
Example application (2)

Simulated LTE cell

- Link layer LTE model
- 5 MHz bandwidth
- K = 10 CIF streaming users
- Radio resources allocated using utility curves as:

$$\arg\max_{(\alpha_1,\dots,\alpha_K)} \sum_{k=1}^{K} U_k(\alpha_k)$$

$$s.t. \sum \alpha_k = 1$$



[4] A. El Essaili et al. "Quality-of-Experience driven Adaptive HTTP Media Delivery", IEEE ICC 2013

Conclusion

- Novel No-Reference PSNR estimation method
- Only 2 parameters needed for estimation:
  - Bitrate of the video stream
  - Mean QP of the I-frames
- Above 0.95 Pearson correlation and around 1dB RMSE for various encoding configurations.

- Further Work:
  - Reduce the bias with additional information about the video content.
  - Evaluate the effect of different resolutions and different framerates.