First analysis on the large scale dataset

What we can learn by « only » comparing objective measurements

Enrico Masala
Politecnico di Torino, Italy
enrico.masala@polito.it

Contribution to VQEG JEG-Hybrid session in Stockholm, Jul 2014

Context and Aim

- Joint work with
  - Glenn Van Wallendael
  - Marcus Barkowsky

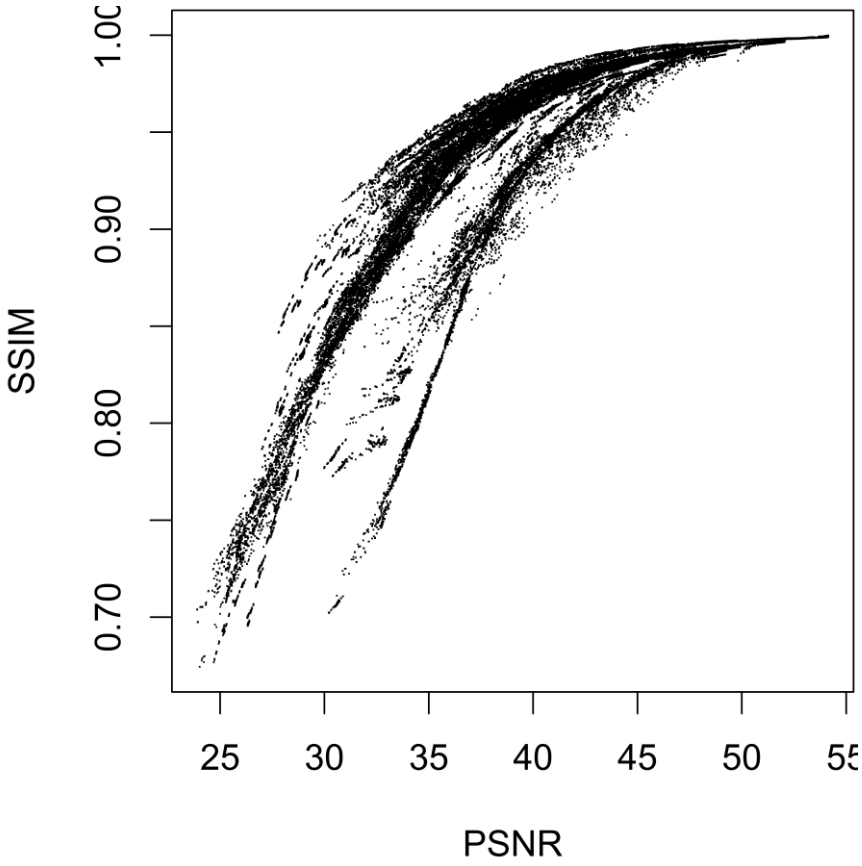
- Large number of objective quality measures available
  - 5 metrics: PSNR, SSIM, VIF, VQM, PVQM
  - 59520 video sequences
  - Parameters available also for each frame of the 250 frames

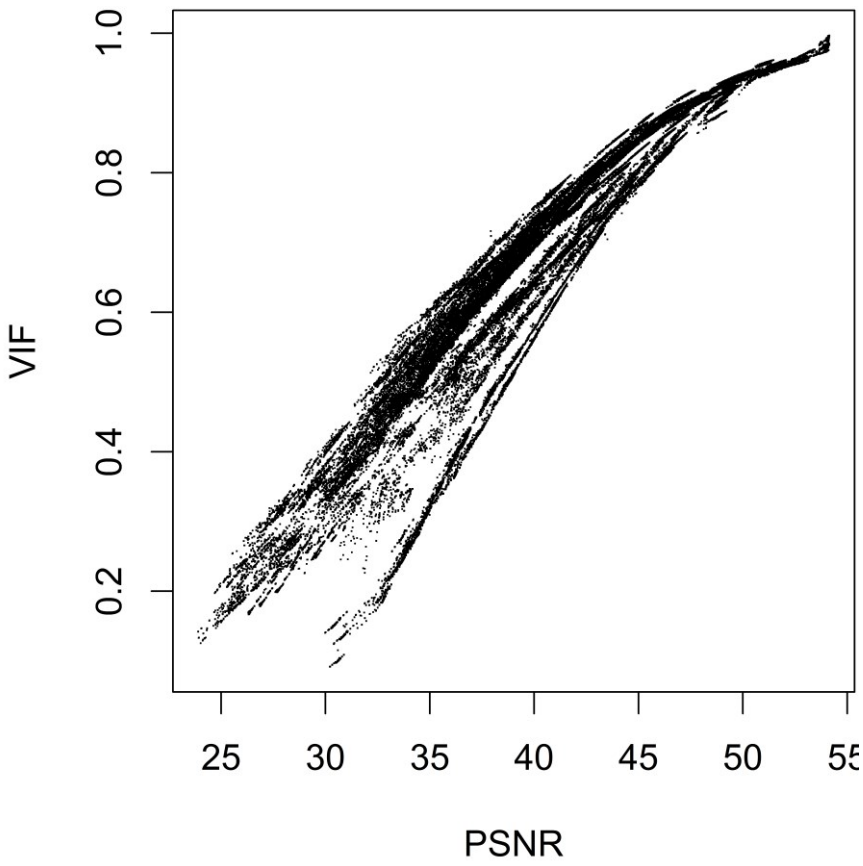
- Is it possible to identify some interesting behavior that deserve further investigation?
  - Potentially leading to better design of hybrid quality metrics

Key Ideas

- Identify cases that look interesting
- Methods
  - Scatterplots
  - Agreement / disagreement between metrics
  - Variance of results
  - Analysis w.r.t. coding parameters
  - …
- Aim: subjective evaluation to be done only on a very limited subset of the database

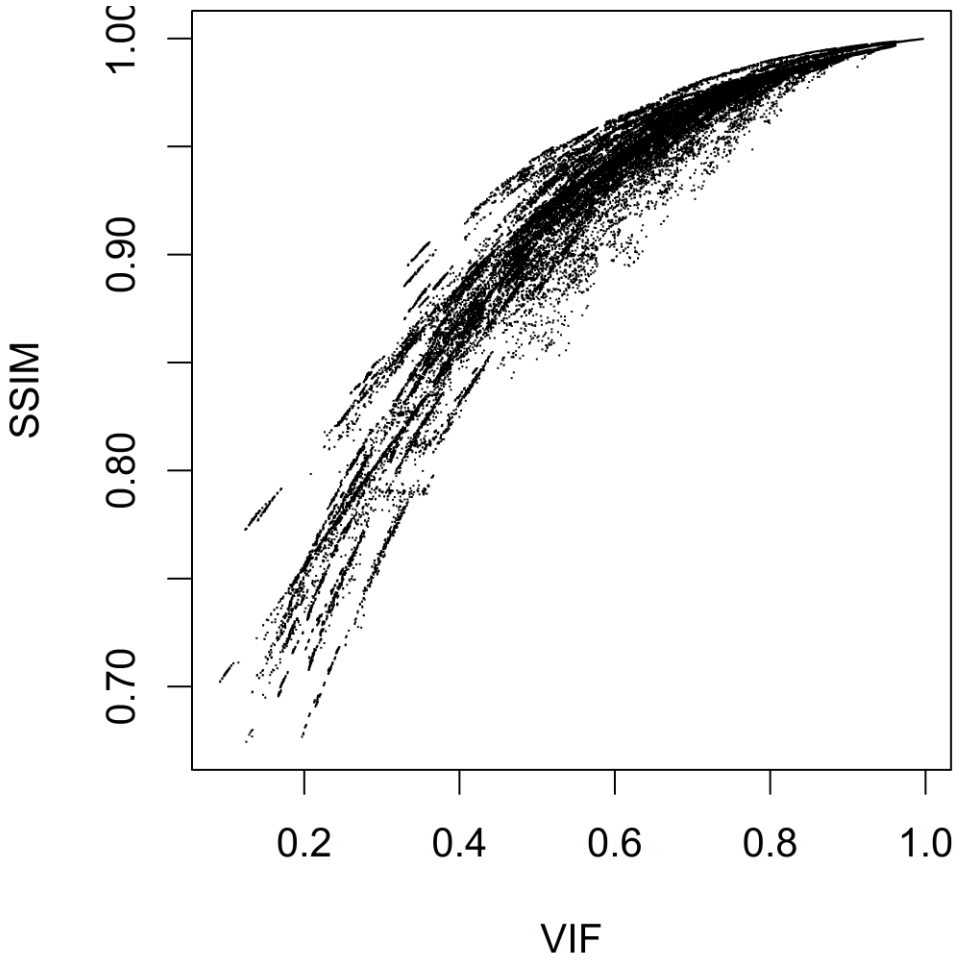
Scatter Plots





PSNR clipped to about 54 dB for identical images

Scatter Plots



Reason of Disagreement

- All pairs of sequences and 3 metrics (PSNR, SSIM, VIF): when they do not agree on which is the better one among the two, which metric is the cause of disagreement?

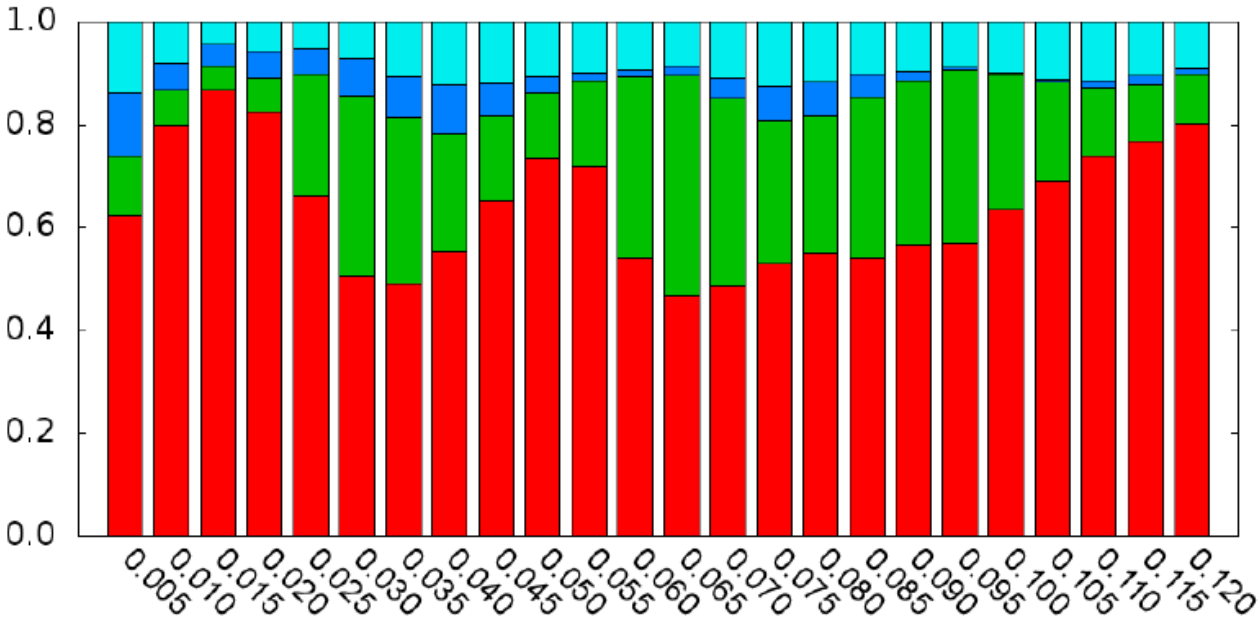
Sequence	Pairs with disagreement	Due to PSNR	Due to SSIM	Due to VIF
src01	3.32%	14.47%	60.72%	24.80%
src02	2.64%	40.74%	45.70%	13.56%
src03	6.27%	61.97%	9.30%	28.73%
src04	4.55%	51.17%	11.76%	37.06%
src05	3.30%	37.89%	18.16%	43.95%
src06	4.99%	28.92%	13.84%	57.24%
src07	6.17%	69.45%	7.41%	23.14%
src08	3.93%	24.58%	59.33%	16.09%
src09	7.65%	20.89%	53.62%	25.49%
src10	3.81%	39.76%	12.55%	47.70%

Reason strongly vary among sequences

PSNR limited to about 54 dB for identical images

Reason of Disagreement

- Linear rescaling of metrics PSNR, SSIM, VIF to [0..1]
- Reason of disagreement as a function of distance (Euclidean)



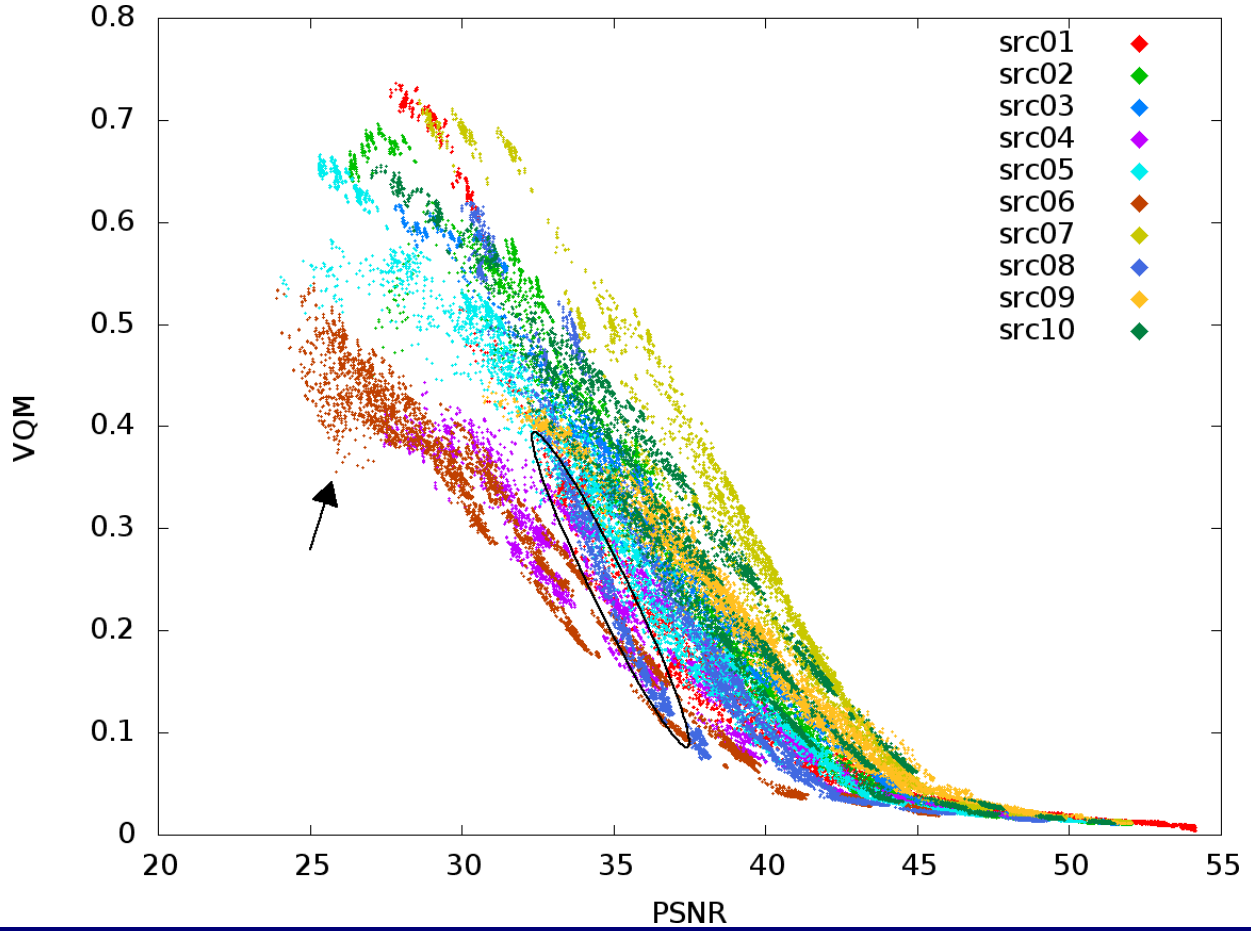
Light blue: VIF
Blue: SSIM
Green: PSNR
Red: agreement

seq03

X axis: Euclidean distance of the three normalized differences (PSNR, SSIM, VIF)

Disagreement between Metrics (per sequence)

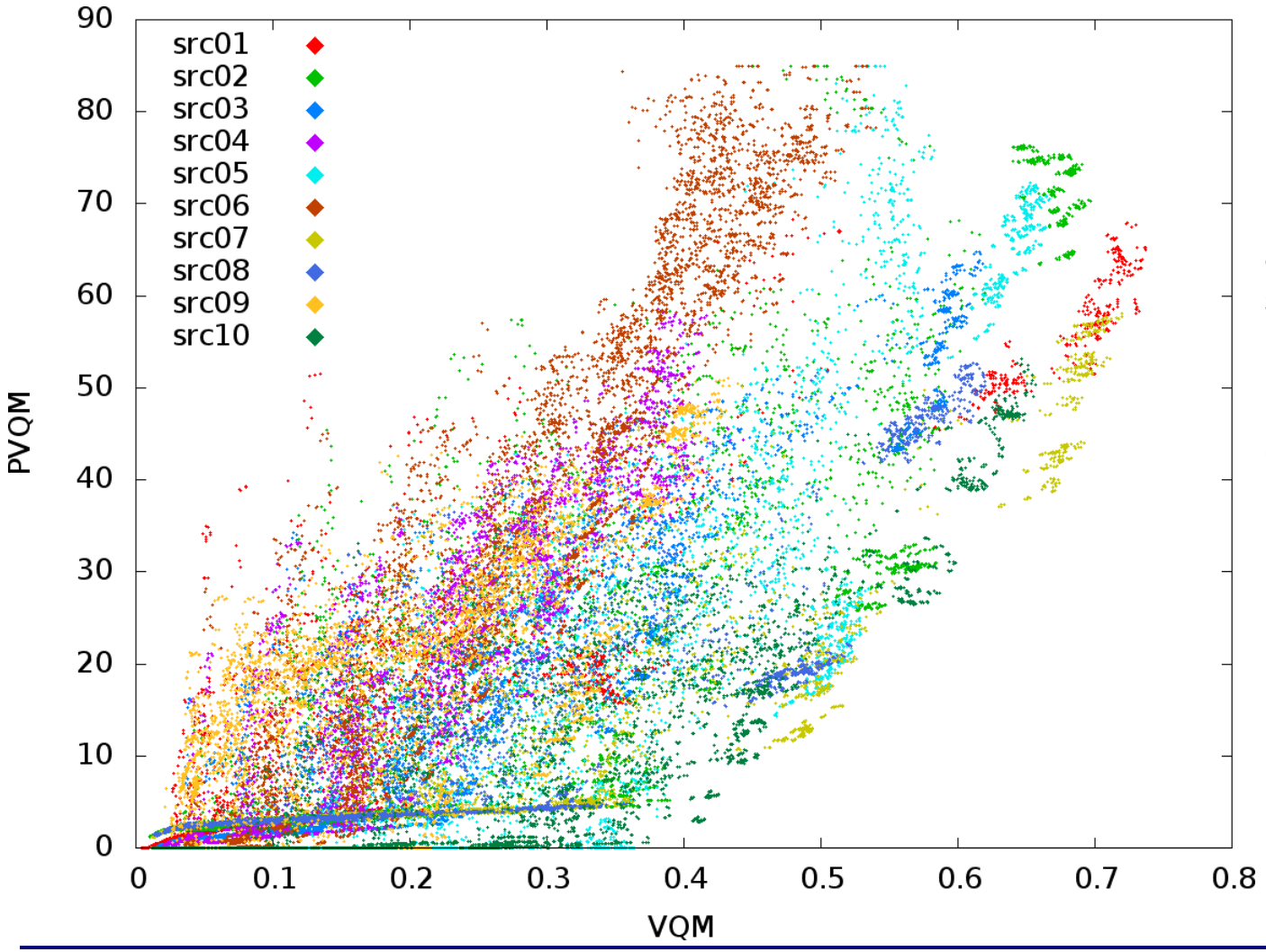
- Situations that need further investigation



Arrow:
High value vs low value

Ellipse:
Strong VQM variation for limited PSNR range (src08)

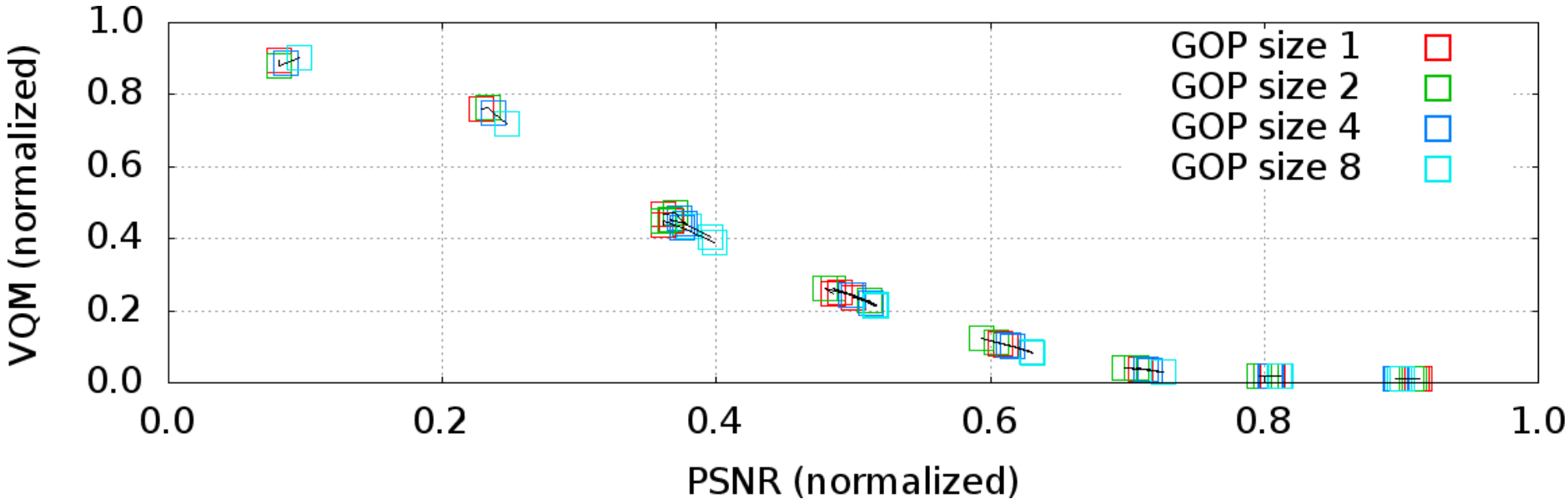
PVQM vs VQM



High correlation: Dense points arranged in a straight line

Low correlation: Other points

Dependency on Coding Parameters



Points are average for same condition except:

- GOP size
- Rate control parameter
  - left side: VBR (fixed) QP
  - right side: CBR (rate controlled)

Conclusions

- Although only objective metrics are available in the database, cases that deserve further investigation can be identified by some statistical analysis
- Subjective experiments could be run for those cases to better understand the root cause of the disagreement
  - Potentially leading to better design of hybrid quality metrics
- Future work will address
  - Sequences with impairments due to data loss
  - Different sequences (different resolutions?)