Equal quality threshold analysis

Marcus Barkowsky



Motivation

- Warning: Preliminary results!
- Random 13000 of the 60000 HEVC sequences analyzed

Setup:

- Forming 13.000² pairs of video sequences
- □ Calculating difference k(x)between PSNR, SSIM, VIFP scores for sequences "A" and "B" by k(x) = s("A") - s("B")
- For each difference score, taking a threshold decision with threshold t:

$$d(x) = \begin{cases} -1 & \text{if } k(x) < -t & \text{"B is better than A"} \\ 0 & \text{if } -t \le k(x) \le t & \text{"B is equal to A"} \\ 1 & \text{if } k(x) > t & \text{"A is better than B"} \end{cases}$$

• Comparing the trinary decisions of the three metrics, i.e.

$$|d_{PSNR}(x) - d_{SSIM}(x)|$$

results in:

- 0 for agreement
- 1 for false tie (weak error)
- 2 for false ordering (strong error)



- 2 -8/7/2014

False ordering errors



False ordering percentage



 t_{VIFP} in percent of maximum difference value



Marcus Barkowsky: Subjective experiment dataset for joint development of HNR algorithms

- 3 -8/7/2014 ... but false tie errors increase

• For
$$t_{PSNR} = 0$$

False tie percentage



Conclusions

- ... is there an optimum?
- If yes, then what is the weighting for the different error cases?

