Equal quality threshold analysis

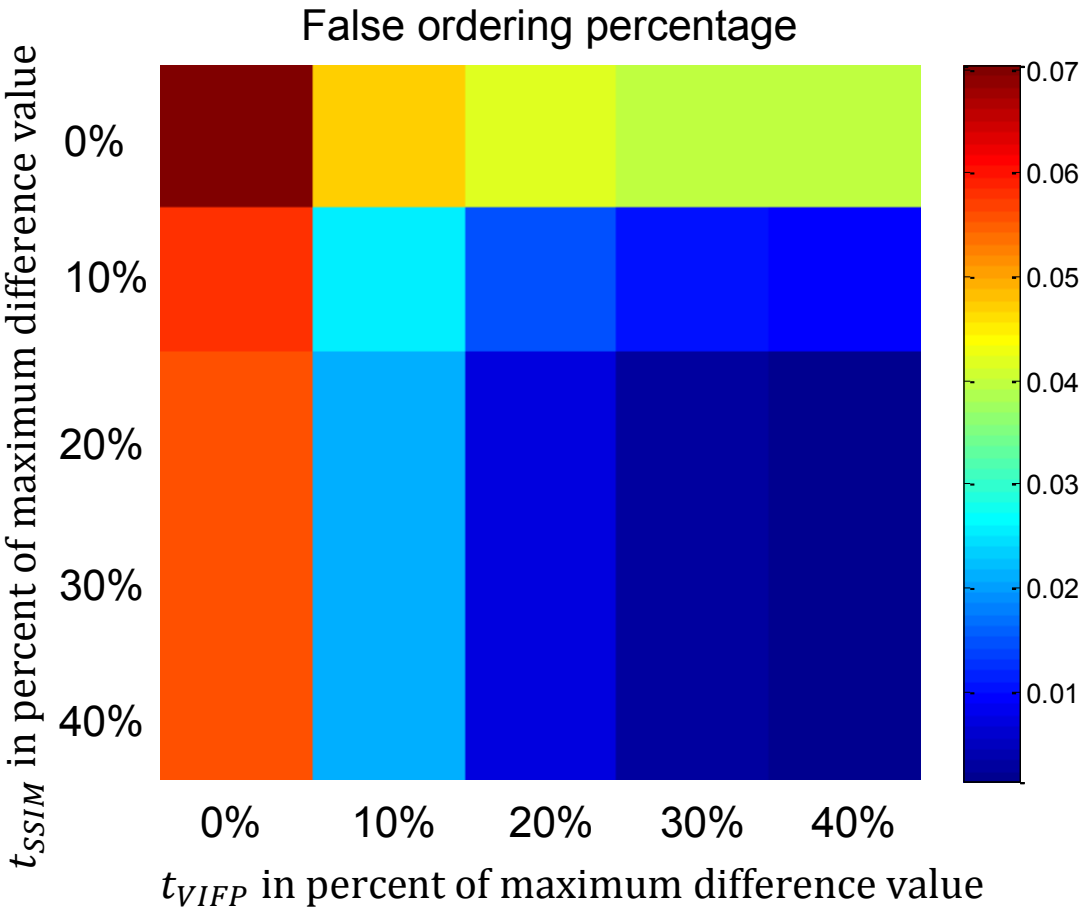
Marcus Barkowsky

Motivation

- Warning: Preliminary results!
- Random 13000 of the 60000 HEVC sequences analyzed
- Setup:
  - Forming 13.000^2 pairs of video sequences
  - Calculating difference *k(x)*
    between PSNR, SSIM, VIFP scores for sequences “A” and “B” by
    $$k(x) = s(\text{"A"}) - s(\text{"B"})$$
  - For each difference score, taking a threshold decision with threshold t:
    $$d(x) = \begin{cases} -1 & if\ k(x) < -t \quad "B\ is\ better\ than\ A" \\ 0 & if\ -t \le k(x) \le t \quad "B\ is\ equal\ to\ A" \\ 1 & if\ k(x) > t \quad "A\ is\ better\ than\ B" \end{cases}$$
  - 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)

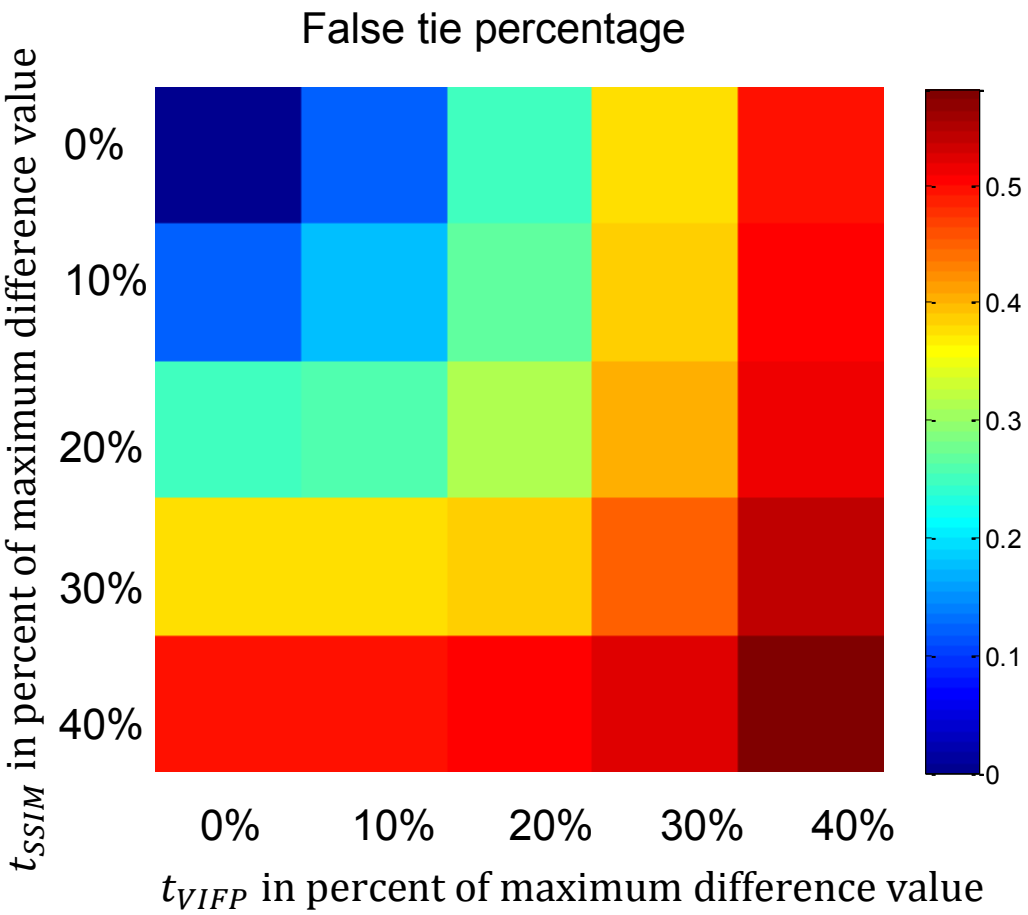
False ordering errors

- For $t_{PSNR} = 0$



… but false tie errors increase

- For $t_{PSNR} = 0$



Conclusions

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