

### FROM LARGE-SCALE TO SMALL-SCALE DATABASE

<u>Ahmed Aldahdooh</u>, Enrico Masala, Glenn Van Wallendael, Marcus Barkowsky, and Patrick Le Callet

VQEG meeting, May 2017

#### **OBJECTIVE**

Identify significant HRCs for:

- Subjective experiments
- Machine-learning-based VQA

How combine datasets? How to select contents? How to select HRCs? How to evaluate your HRCs subsets?

Introduction to "Improved Performance Measures for Learning-based Video Quality Assessment Algorithms"

#### LARGE-SCALE DATABASE

Different correlation scores may be obtained when testing an objective video quality (VQ) measurement using two different databases (and cannot really be averaged)

- Lack of content variety in the databases.
- Lack of different HRCs in the experiments.

Go for Large-scale?

- To evaluate objective measurements that is difficult to achieve in subjective assessment due to Limited HRCs.
  - Agreement of objective measures.
  - Not convenient for frame-based analysis? Consistency within a video sequence.
  - The impact of source contents and the encoder parameters are studied.

### SMALL-SCALE DATABASE

#### After Analysis: go for small-scale?

- Identify significant HRCs for:
  - Subjective experiments
  - Machine-learning-based VQA

#### What we need is to choose HRCs that cover a good variety of targets.

#### FIRST PART: HRC SELECTION ALGORITHMS

# QUALITY/BITRATE-DRIVEN HRCS SUBSET CONTENT-DRIVEN HRCS SUBSET





#### **OBSERVATION**



#### $\mathsf{RESULTS}-1$



#### RESULTS — 2 - STD. OF THE RANKS MAGNITUDE



#### SECOND PART: IMPROVED PERFORMANCE MEASURES

After we have these subset, how do they perform?



#### **EXPERIMENTS STEPS**

To test the goodness of the elected HRCs subsets.

Not to evaluate the prediction models



11

#### SHORTCOMING WITH PLCC AND RMSE

0

Measured VQM

0

Measured VQM

#### Random 1 subset HRC3 Content-based subset HRC1 Random 2 subset HRC4 Quality/bitrate-based subset HRC2 Random 3 subset HRC5 PCC=0.98006, RMSE=0.028603 PCC=0.9728, RMSE=0.035566 PCC=0.9742, RMSE=0.036299 PCC=0.97515, RMSE=0.043059 PCC=0.97745, RMSE=0.036 Content-based subset HRC1 ited VQM VQM ğ Ś Q Fitted for test Fitted model 0.5 ge 0.5 g 0.5 0.5 0.2 0.6 0.8 0.6 0.8 0.6 0.8 0.2 0.4 0.6 0.8 0.6 0.4 0.2 0.4 0.2 0.4 0.2 0.4 0.8 0 Ω Measured VQM Measured VQM Measured VQM Measured VQM Measured VQM PCC=0.98121, RMSE=0.037537 PCC=0.9715, RMSE=0.034228 PCC=0.98218, RMSE=0.028684 PCC=0.97866, RMSE=0.033037 Quality/bitrate-bas subset HRC2 PCC=0.98199, RMSE=0.032191 oted VQM VQM Nov Ś Ś 0.5 8 0.5 0.5 g 20.5 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0 Measured VQM Measured VQM Measured VQM Measured VQM Measured VQM PCC=0.96494, RMSE=0.039446 PCC=0.97412, RMSE=0.035327 PCC=0.98456, RMSE=0.029213 PCC=0.9842, RMSE=0.039315 PCC=0.98016, RMSE=0.035738 Predicted VQM VQM Trained on ğ ğ Q odicted 0.5 0.5 te 0.5 E e 0.5 ñ 0.8 0.2 0.4 0.6 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0 1 0 Measured VQM Measured VQM Measured VQM Measured VQM Measured VQM PCC=0.95696. RMSE=0.042024 PCC=0.9709. RMSE=0.037648 PCC=0.98048, RMSE=0.032568 PCC=0.99035, RMSE=0.027797 PCC=0.98349, RMSE=0.032657 Random 2 subset HRC4 Predicted VQM VQM ğ ģ Ś 0.5 ted 0.5 e 0.5 e 0.5 0.6 0.8 0.6 0.6 0.8 0.6 0.2 0.4 0.2 0.4 0.8 0.2 0.4 0.2 0.4 0.8 0.2 0.4 0.6 0.8 0 0 Measured VQM Measured VQM Measured VQM Measured VQM Measured VQM PCC=0.96762, RMSE=0.036266 PCC=0.97598, RMSE=0.033203 PCC=0.97693, RMSE=0.034602 PCC=0.98384, RMSE=0.035121 PCC=0.98849, RMSE=0.025857 Random 3 subset HRC5 edicted VQM VQM ğ ğ ğ 0.5 g 0.5 0.5 0 ! 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8 0.2 0.4 0.6 0.8

0

Measured VQM

Evaluation on

**RMSE & Correlation** cannot tell us exactly which HRC set is better

> AHMED ALDAHDOOH: FROM LARGE-SCALE TO SMALL-SCALE DATABASE 12

Measured VQM

0

Bitstream-based NR VQA

0

Measured VQM

#### SHORTCOMING WITH PLCC AND RMSE

Another Model RMSE and Correlation can tell us something



AHMED ALDAHDOOH: FROM LARGE-SCALE TO SMALL-SCALE DATABASE 13

### (1) RESIDUAL ANALYSIS USING PCA

PCs

PCs

**Trained** on

#### Evaluation on Random 1 subset HRC3 Content-based subset HRC1 Quality/bitrate-based subset HRC2 Random 2 subset HRC4 Random 3 subset HRC5 Content-based subset HRC1 $P_{RPCA_V}^{SRC} = 70.9695$ $P^{SRC}_{RPCA\_V} = 74.7788$ $P_{RPCA_T}^{SRC} = 55.5425$ $P_{RPCA_V}^{SRC} = 76.7315$ $P_{RPCA_V}^{SRC} = 79.8881$ 100 % 50 Var. Quality/bitrate-based subset HRC2 1 2 3 4 5 6 7 8 9 2 3 4 7 8 1 2 3 4 5 7 8 9 1 2 3 4 5 6 2 3 4 $\begin{array}{c} \mathsf{PCs} \\ P^{SRC}_{RPCA,T} = 48.9377 \end{array}$ $\begin{array}{c} {\sf PCs} \\ P^{SRC}_{RPCA\_V} = 61.298 \end{array}$ $\begin{array}{c} \text{PCs} \\ P_{RPCA\_V}^{SRC} = 58.1612 \end{array}$ $\begin{array}{c} \mathsf{PCs} \\ P^{SRC}_{RPCA\_V} = 65.9828 \end{array}$ PCs $P^{SRC}_{RPCA_V} = 55.0041$ 100 1 2 3 4 5 1 2 3 4 5 1 2 3 4 1 2 3 1 2 3 4 5 $\begin{array}{c} \mathsf{PCs} \\ P^{SRC}_{RPCA\_V} = 76.1369 \end{array}$ $\begin{array}{c} \mathsf{PCs} \\ P^{SRC}_{RPCA\_V} = 62.5504 \end{array}$ $\begin{array}{c} \text{PCs} \\ P_{RPCA\_T}^{SRC} = 53.1866 \end{array}$ $\begin{array}{c} \mathsf{PCs} \\ P^{SRC}_{RPCA\_V} = 68.1911 \end{array}$ $\begin{array}{c} \mathsf{PCs} \\ P^{SRC}_{RPCA\_V} = 83.1468 \end{array}$ Random 1 subset HRC3 % 50 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 7 8 9 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 6 $\begin{array}{c} \text{PCs} \\ P^{SRC}_{RPCA\_V} = 57.4807 \end{array}$ $\begin{array}{c} \text{PCs} \\ P_{RPCA_V}^{SRC} = 54.884 \end{array}$ PCs PCs PCs $P_{RPCA_V}^{SRC} = 84.8056$ $P^{SRC}_{RPCA\_T} = 53.781$ $P_{RPCA_V}^{SRC} = 58.1597$ Random 2 subset HRC4 100 100 % Var. 1 2 3 4 5 1 2 3 4 5 6 1 2 3 4 5 6 1 2 3 4 5 6 6 7 8 9 1 2 3 4 5 6 $\begin{array}{c} \mathsf{PCs} \\ P^{SRC}_{RPCA\_V} = 97.8162 \end{array}$ $\begin{array}{c} \mathsf{PCs} \\ P^{SRC}_{RPCA\_V} = 97.9896 \end{array}$ PCs PCs PCs $P_{RPCA-V}^{SRC} = 97.7432$ $P^{SRC}_{RPCA_V} = 94.8281$ $P^{SRC}_{RPCA,T} = 53.8244$ Random 3 subset HRC5 % % 50 50 3 4 5 6 7 8 9 1 2 3 4 5 8 1 2 3 4 5 6 7 8 9 1 2 3 4 5 6 3 4 5 6 6

PCs

How residual structured?

Any sign for systematic redundancy in the residual?

AHMED ALDAHDOOH: FROM LARGE-SCALE TO SMALL-SCALE DATABASE 14

PCs

#### Pixel-based NR VQA

PCs

# (2) CONFIDENCE INTERVALS OF PREDICTED DATA

Pixel-based NR VQA

How much of the predicted data lies within Cl of the trained model?

#### **Remember:**

Black lines: CI boundaries of the predicted data of the trained model. Red Lines: CI boundaries of the predicted data of the validation data



AHMED ALDAHDOOH: FROM LARGE-SCALE TO SMALL-SCALE DATABASE 15

# (3) CONFIDENCE INTERVALS OF TRAINED MODELS

Is the model stable when validation data is used?

#### **Remember:**

Black lines: CI boundaries of the model coefficient of when training data is used. Red Lines: CI boundaries of the model coefficient when validation data is used.



# (4) INTERSECTION ANALYSIS

	Case	Icon	Condition	Note			
$G = \frac{i}{\max(b,r)^2}$	1		b = r = i	Typical case for validating on the training data, this is considered the perfect fitting, i.e. all three areas are identical. Refer for example to the main diagonal $X(n,n)$ in Fig. 6. In this case, $G = \frac{1}{\max(b,r)}$ . To compare between different models or data, the lower the $\max(b,r)$ , i.e. the smaller the larger CI, the better.			
The interaction between black lines and red lines?	2		r = i	The validation data is better predicted than the training data and the CI lie completely within the boundaries of the trained model. This is likely to be a default of			
The higher the overlap the better.				the validation data and thus reduces the goodness as compared to Case 1. In this case, $G = \frac{r}{b^2}$ .			
	3		b = i	The validation data is less well predicted than the training data but the validation CI covers completely the training CI. This is considered a case of overfitting of the model and should thus be penalized compared to case 1. In this case, $G = \frac{b}{r^2}$ .			

# (4) INTERSECTION ANALYSIS (CONT.)



# (4) INTERSECTION ANALYSIS (CONT.)



#### PERFORMANCE MEASURE COMPARISON

Performance measure	Pixel-based NR VQA (Proposed)					Bit-stream-based NR VQA				
	Content	RD	Rand 1	Rand 2	Rand 3	Content	RD	Rand 1	Rand 2	Rand 3
PLCC Cross-dataset	3	1	4	2	5	4	1	3	5	2
PLCC Leave-one-out	2	1	5	3	4	2	4	3	5	1
PLCC Challenging HRCs	2	1	3	4	5	1	2	3	5	4
RMSE Cross-dataset	3	1	4	2	5	5	1	4	3	2
RMSE Leave-one-out	2	1	4	3	5	3	5	4	1	2
RMSE Challenging HRCs	1	2	3	4	5	1	3	5	2	4
$P_{\text{RPCA_T}}^{\text{SRC}}(\frac{n}{m},m), P_{\text{RPCA_V}}^{\text{SRC}}(\frac{n}{m},m)$	3	1	4	2	5	1	2	1	1	3
$P_{\text{DCL}V}(\delta, n) = \frac{i}{\rho}$	3	1	4	2	5	2	1	3	5	4
P <sup>(b,r,i)</sup> <sub>GModel</sub>	3	1	4	2	5	2	1	3	4	5
P <sup>(b,r,i)</sup> <sub>GData</sub>	3	1	4	2	5	5	1	2	3	4
Average	2.5	1.1	3.9	2.6	4.9	2.6	2.1	3.10	3.4	3.10

#### THANKS!

Questions