

A Novel Approach to Video Quality Assessment using a Quality Ruler Method

Pedro G. Freitas Judith A. Redi Mylène C.Q. Farias

University of Brasília

TU Delft

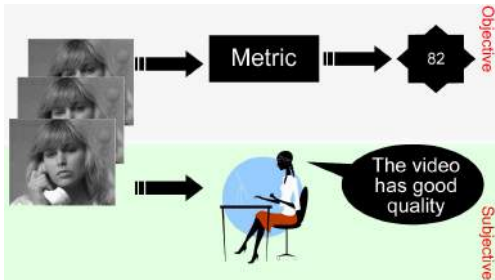
Friday 12th June, 2015



Overview

- 1 Introduction
- 2 Subjective Experiments
- 3 Proposed Methodology
- 4 Results
- 5 Conclusion
- 6 Conclusion

Quality Metrics



Quality metrics are the gathering and/or use of values that indicate how accurately a system or service can reproduce video or perform actions within desired levels.

- **Objective Quality** is the determination of accuracy or the ability of a system to provide desired results using evaluation criteria and sources that are repeatable.
- **Subjective Quality** is determination of accuracy or the ability of a system to provide desired results using evaluation sources that can vary (such as the opinions of people).

Why experiments are needed?

Controlling Quality



Subjective
(MOS)



Objective
(MOSp)

*The mean opinion score (MOS) provides a numerical measure of the quality of human perceived information.

$$MOS_i = \frac{1}{n} \sum_{j=1}^n o_{ij}$$

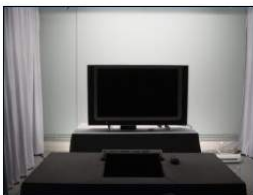
How to measure quality subjective assessment?

How to extract the subjective data?

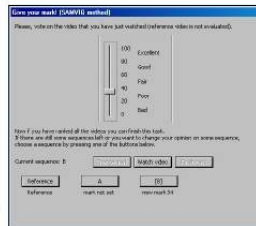
Getting a mean human quality evaluation



Observers
(raters)



Environment
(Experimental Setup)



Methodology

Methodology Selection Criteria

- Measurements reliability
- Confidence in the comparison of different experiments
- Standardization
- Confusion in the test set
- Required effort per sample
- Implementation easiness
- Result analysis easiness
- Psychophysical significance of the results

Methodology Selection Criteria

For **images**, we know that...

	Single Stimulus (SS)	Quality Ruler
Reliability	Medium	High
Confidence	Medium	High
Standardization	ITU-R BT.500-11	ISO 20462-3
Confusion in the test set	Low	Low
Required effort per sample	Low	Medium
Implementation easiness	High	Low
Result analysis easiness	High	High
Significance of results	Low	High

- Rasmussen, D. Ren, et al. **“ISO 19751 macro-uniformity.”** Electronic Imaging 2006. International Society for Optics and Photonics, 2006.
- Redi, Judith, et al. **“Comparing subjective image quality measurement methods for the creation of public databases”.** IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2010.

Quality Ruler: Research Question 1

Can Quality Ruler be used as methodology to assess videos?



Quality Ruler: Research Question 2

If yes, is it a good approach?



Single Stimulus: Interface



Previous

Score

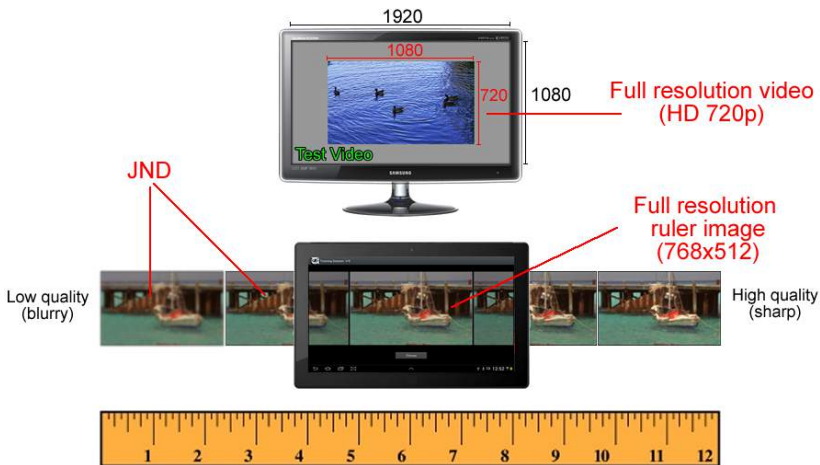
Next

0 1 2 3 4 5 6 7 8 9 10

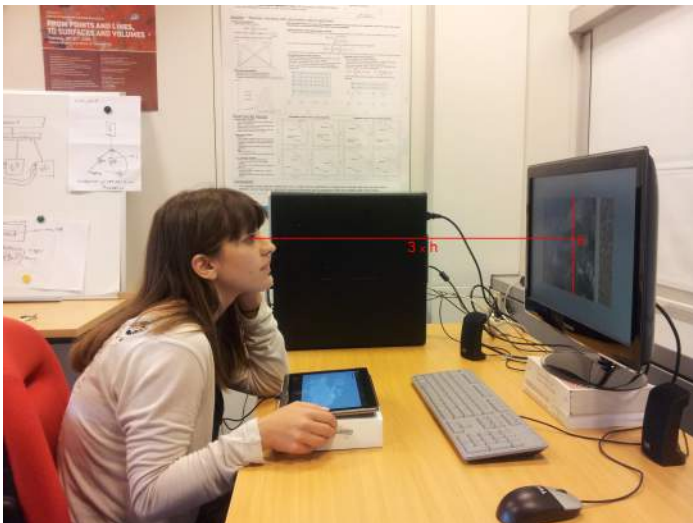
Imperceptible

Very Annoying

Proposed Methodology: Interface



Proposed Methodology: Setup



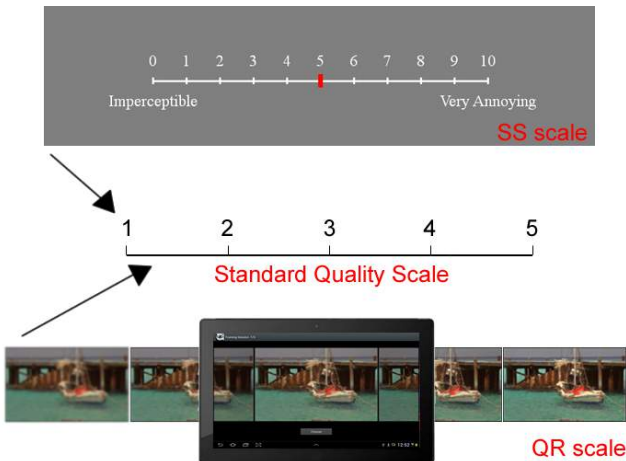
Proposed Methodology: Setup



- 7 original videos (scenes)
- 3 levels of artifacts
 - Blockiness
 - Blurring
 - Packet loss
- 49 videos in total
- Number of participants
 - Single Stimulus: 24
 - Quality Ruler: 17
- Same video sequences for both SS and QR
- Same environmental settings (lights, monitor screen, distance, etc)

Analysis: Putting on a common scale

For comparison, the data of both experiment was scaled to a common continuous 5-point scale.

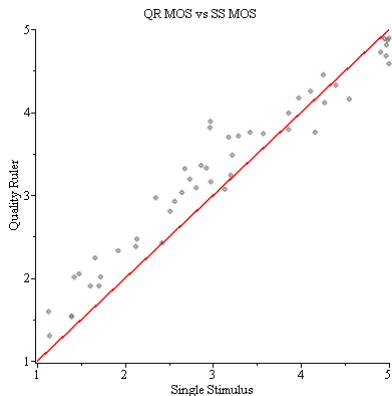


Analysis: Objective analysis

The analysis was divided into two parts:

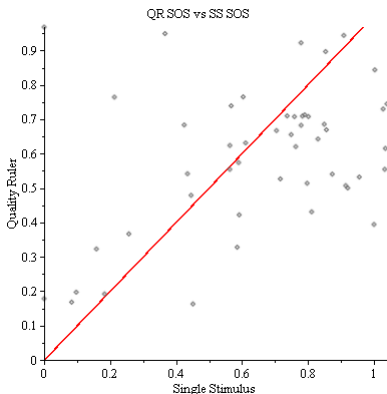
- ① Parallel form reliability analysis. It was used to answer the first question (*Can quality ruler be used as methodology to assess videos?*)
- ② Inter-rater reliability and internal consistency analysis. It was used to answer the second question (*What is the better approach?*)

Results 1: Experiment Comparison



- Linear Correlation: 0.9663
- Spearman's Correlation: 0.9643
- Kendall's Correlation: 0.8511
- RMSE: 0.3871
- Outliers Ratio
($[MOS - 2\sigma, MOS + 2\sigma]$): 0

Results 2: Internal Consistency (Cronbach's alpha)



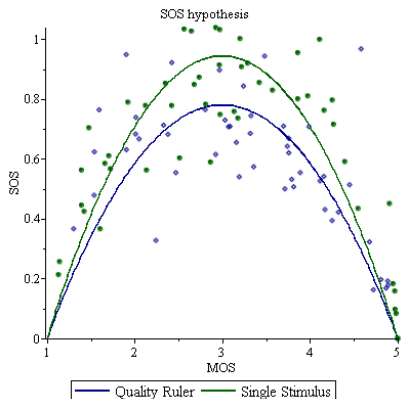
- OBS: Low correlation between **Standard deviation of Opinion Scores (SOS)**

$$SOS_i = \sigma(o_{ij})$$

- Cronbach's alpha for QR: **0.9810159**
- Cronbach's alpha for SS: **0.9867592**

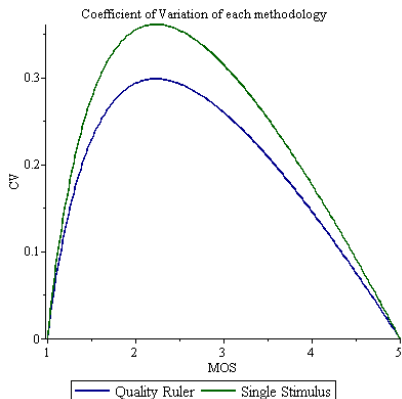
Cronbach's alpha	Internal consistency
$\alpha \geq 0.9$	Excellent (High-Stakes testing)
$0.7 \leq \alpha < 0.9$	Good (Low-Stakes testing)
$0.6 \leq \alpha < 0.7$	Acceptable
$0.5 \leq \alpha < 0.6$	Poor
$\alpha < 0.5$	Unacceptable

Results 2: Internal Consistency (Hossfeld's alpha)



- $SOS(x)^2 = \alpha (-x^2 + 6x - 5)$
- Hossfeld's alpha for QR:
0.195065739659081
- Hossfeld's alpha for SS:
0.23603115471979

Results 2: Internal Consistency (Coefficient of Variation)

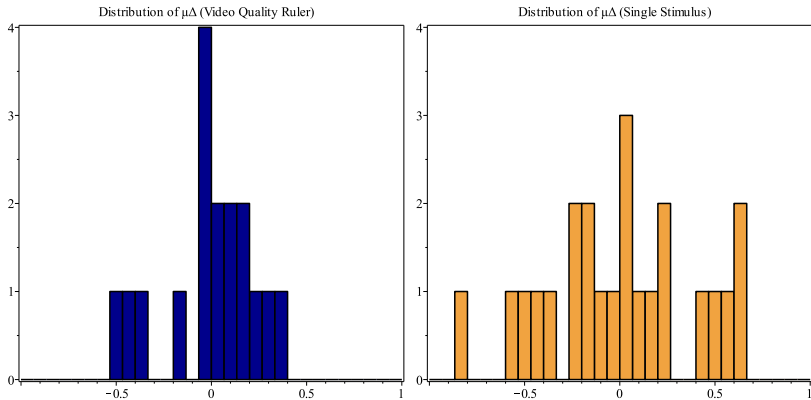


- Maximum coefficient of variation for QR: **0.2980339300**
- Average coefficient of variation for QR: **0.1927644730**
- Maximum coefficient of variation for SS: **0.3606235147**
- Average coefficient of variation for SS: **0.2332466030**

*The coefficient of variation represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one series to another. **The smaller the coefficient, the higher the reliability of mean.**

$$CV(x) = \frac{SOS(x)}{x}$$

Results 2: Internal Consistency (Subject Bias)



Conclusions and Future Work

- Quality ruler method has been adapted for assessment of video quality
- The method was compared with the well-known Single Stimulus methodology
- Can quality ruler be used as methodology to assess videos?
YES
- Is it a good approach? **YES, better than Single Stimulus at least.**
- Future works include the comparison between the quality ruler and other standardized methodologies (double stimulus continuous quality scale (DSCQS) and double stimulus comparison scale (DSCS), pair comparison (PC), etc).

Conclusions and Future Work

For more information:

- P. G. Freitas, J.A. Redi, M. C. Q. Farias and A. F. Silva, “Video Quality Ruler: A New Experimental Methodology for Assessing Video Quality,” in Quality of Multimedia Experience (QoMEX), Seventh International Workshop on, May 2015.

Conclusions

Thank you.

Questions?

Pedro Garcia Freitas

P.GarciaFreitas@tudelft.nl or sawp@sawp.com.br

