artemis
Advanced Research & TEchniques for Multidimensional Imaging Systems

Issues with Video Quality Assessment of Cloud Games with 3D computer graphics

Video Quality Experts Group
July 8-12, 2013 - Meeting
Ghent, Belgium

Rahul Gaurav
ARTEMIS
Télécom SudParis, France

Plan of Talk

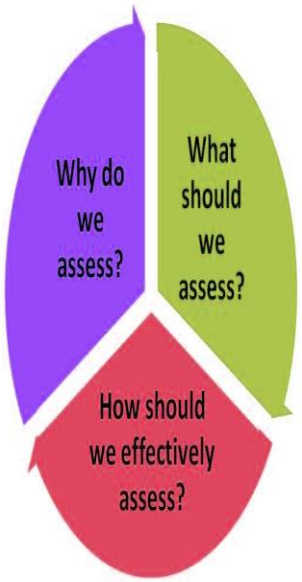
- Introduction
- Project Objective
- Subjective Quality Assessment Experiment
- State of Art Objective Models
- Concluding Remarks
- Future Work

Introduction

- The quality assessment of digital videos has acquired a lot of attention in the video processing community.
- Lot of metrics have been formulated for the purpose of assessing the 2D video quality.
- The correlations can be measured by the opinion scores based on observations of human subjects.
- Apparently, there is a lack of literature on Video Quality Assessment (VQA) of video games or other aspects of 3D computer graphics.

Objective of the Project

- Currently, we are trying to conduct a fully-fledged subjective test strictly following the benchmarks of the ITU.
- Phase 1 : Firstly, the subjective test will be conducted in a similar manner as in case of 2D digital videos to get the idea of user assessment while the gaming videos are broadcasted and the subjects are not playing.
- Phase 2 : This phase covers the complicated task, that is to conduct the subjective VQA in real time when the subjects are asked to assess the quality while playing the game, that is, subjects would be gamers.



Objective (contd.)

- The ultimate motive is to obtain best correlation between the objective and subjective metrics of the compressed video sequences.

Subjective Quality Assessment Experiment

- Viewing Conditions
- Selection of test materials
- Observers selection
- Video session
- Useful information for assessment
- Video quality methods

Viewing Conditions

Parameters	Settings
Peak luminance of the screen	150 cd/m2
Other room illumination	quite low
Height of image on screen (H)	11 cm
Viewing Distance	88 cm
Number of Subjects	30
Maximum Duration of a Session	25 min

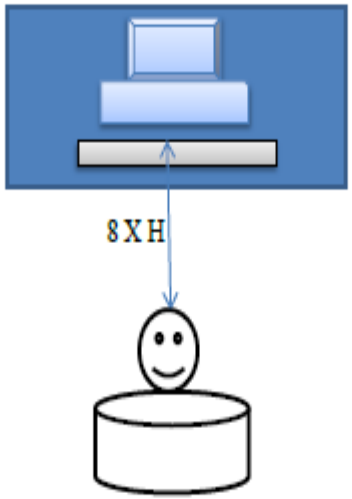
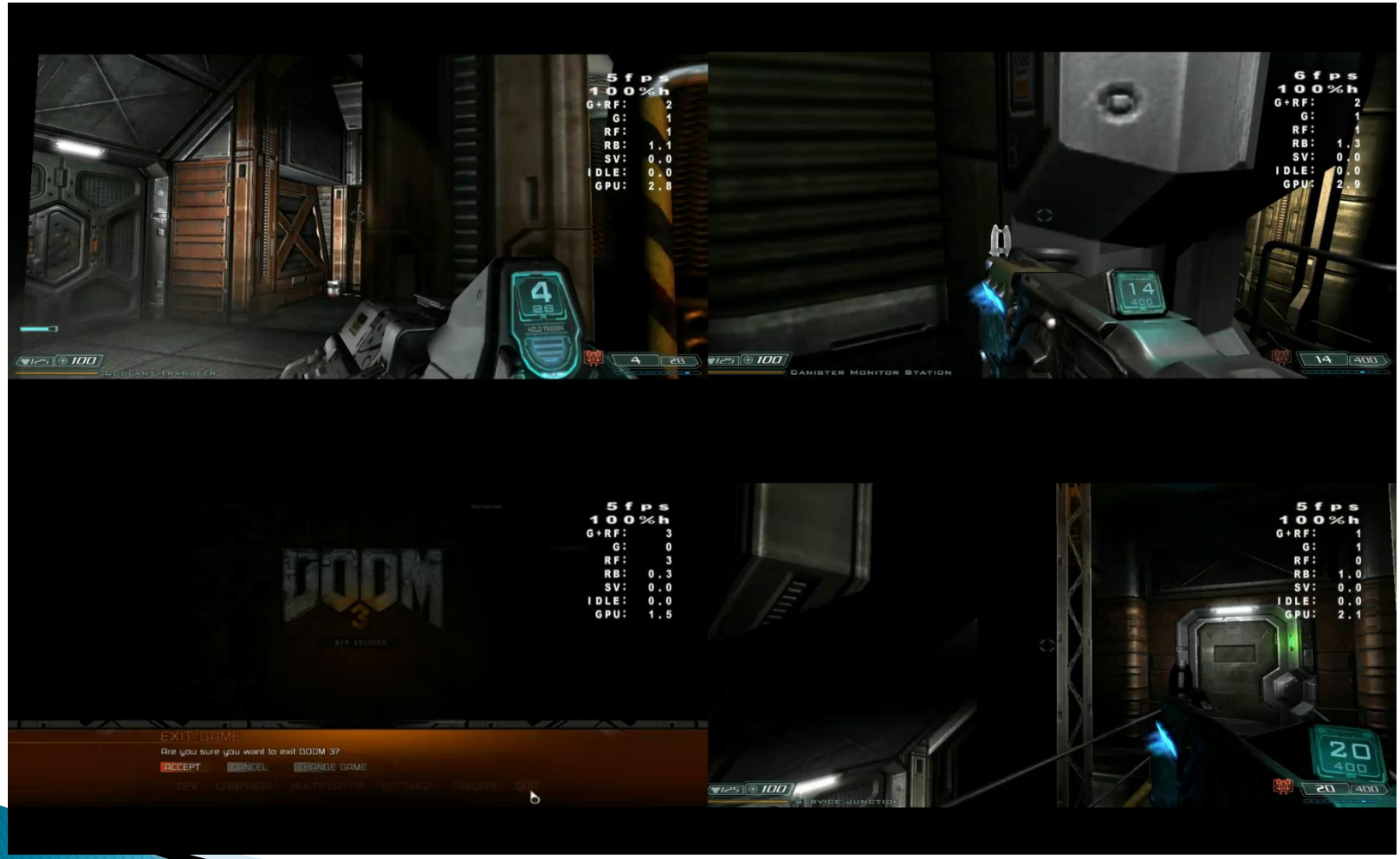


photo courtesy: EPFL Lausanne, Switzerland

Video Samples



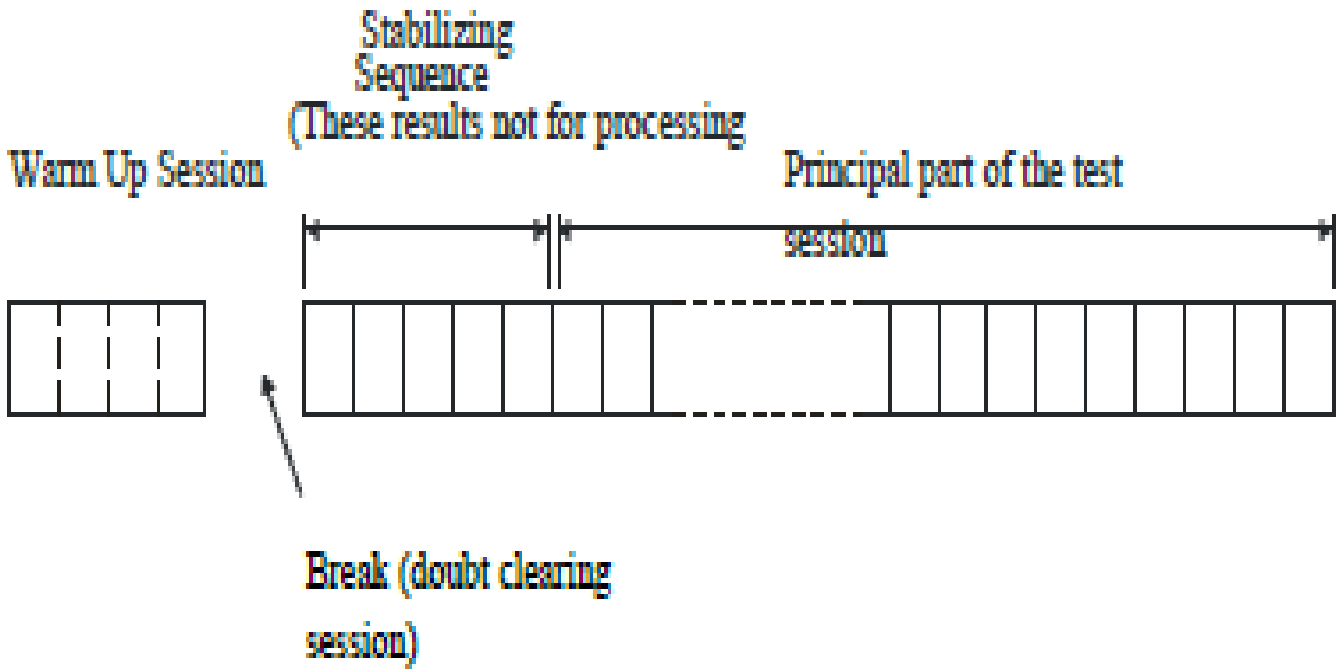
Subjective Testing Design

- A carefully selected random playlist will be prepared comprising of around 24 videos in total, that is, 10 seconds samples produced from the available games with the distorted counterparts.
- Latency and frame drops are also to be dealt with utter seriousness so that while rating, the user does not get bothered by any kind of other issues apart from video distortion based issues.

Observers selection and Training

- Most of the subjects are going to be non-experts in the field of Image/Video Processing in the first phase and avid gamers in the second phase.
- Each video will be rated by 30 subjects in total.
- Each subject will be told about the aim of the experiment one by one and will be carefully instructed about the methods of grading the video sequences.
- Subjects will be given a warm up session in order to familiarise themselves with the environment and experiments before they started the main session.
- They will also be asked to provide ratings for the warm up session to help them understand the testing procedure, these warm up ratings were not used in finding the final Mean Opinion Score (MOS).

Test session presentation structure

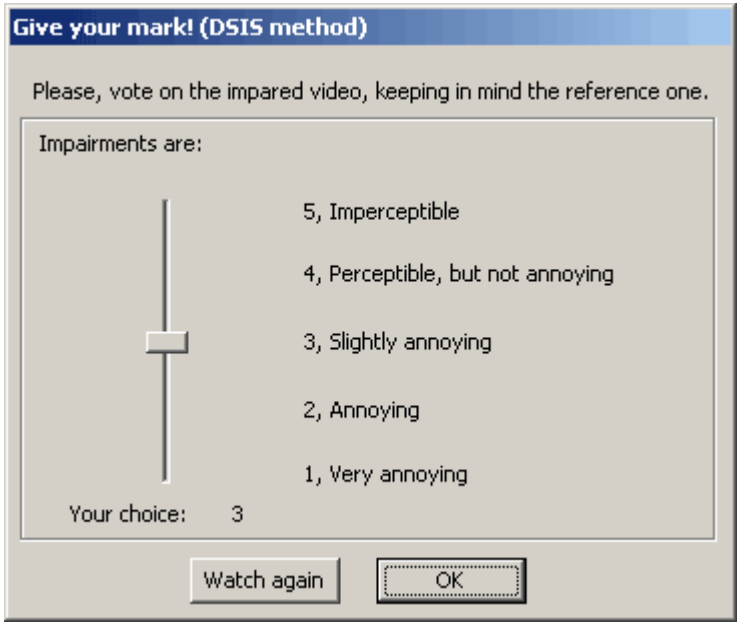


Video Quality Assessment Methods

- The Double Stimulus Impairment Scale (DSIS) tool will be used as it helps us for such cases where we can easily see the distortions produced by the encoding process like blockiness, ringing and blurring.
- This method is useful for assuring the similarities of test condition with respect to the reference condition and it is effective for high quality system evaluation in the multimedia communications.
- For the trial of DSIS method, the observer will be demonstrated with video sequences in pairs where the first one is called the reference sequence and the second one is called the test or impaired sequence.

Grading Scales

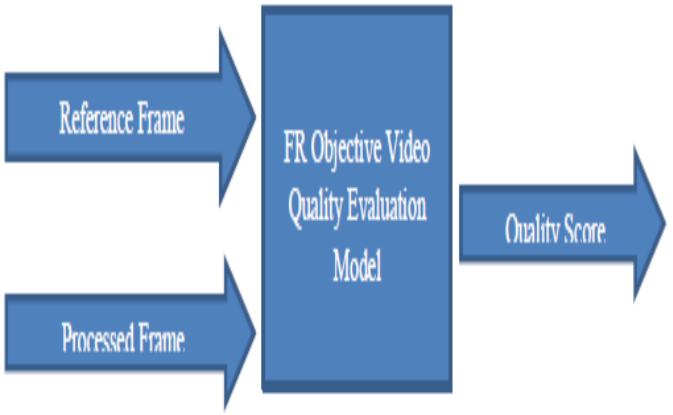
- Using the DSIS methodology, the subjects were asked to rank the impairment in a five grade impairment scale :
- 5: imperceptible
- 4: perceptible, but not annoying
- 3: slightly annoying
- 2: annoying
- 1: very annoying



Video Quality Evaluation Program Interface

Objective Quality Assessment

- We plan to carry out the full reference objective quality assessment test for our selected set of videos.
- We will use various algorithms for this purpose, especially the ones which are popular in the VQA field of study.
- We can see a basic structure of how metrics work in full reference quality evaluation manner.



State of Art Objective Models

- Peak Signal to Noise Ratio (PSNR)
- Structural SIMilarity Gaussian (SSIM G)
- Structural SIMilarity Box Filter (SSIM B)
- Multi-scale Structural SIMilarity Gaussian (MS-SSIM G)
- Multi-scale Structural SIMilarity Box Filter (MS-SSIM B)

Simulation Results and Discussion

- Processing of Subjective Quality Assessment Scores
- Processing of Objective Quality Assessment results
- Metrics Performance
- Discussion of Results

Calculation of Mean Scores

Score of observer i for the presentation j is given as u_{ji} and for each and every presentation carried out in the experiment, we find the mean score, given by $\overline{u_j}$

$$\overline{u_j} = \frac{1}{N} \sum_{i=1}^{N} u_{ji}$$

- Using the above equation, we form a mean opinion score databank.
- The obtained MOS is used to judge the performance of various metrics.

Processing of Objective Quality Assessment results

- Objective experiments on the video database were performed
- We obtain the simulation results for various metrics.

Metrics Performance

- Various results from subjective and objective experiments will be obtained in a proper manner.
- We intend to gauge the performance of the metrics effectively.
- In order to do so, metrics performance estimation using correlation coefficients will be used so that we get the accuracy and monotonicity of the objective metrics when comparing with the subjectively obtained mean opinion scores.

Accuracy of Metrics Performance

- It is that attribute of a metric which helps to predict the subjective ratings with minimum average error.
- It is computed by means of Pearson Linear Correlation Co efficient (PLCC).
- With the help of upcoming algorithm, we compute the PLCC

Algorithm of PLCC

1. MOS and the objective scores are given by length N vector pairs x_i and y_i

2. Compute $\bar{x}$ and $\bar{y}$, the means of respective data sets, that is, the MOS and the objective score

3. Ultimately, PLCC is computed as

$$r_p = \frac{\sum(x - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}}$$

Monotonicity of Metrics Performance

- It estimates the association of increase or decrease in the MOS values to the increase or decrease in the objective metrics value, independently of the magnitude of the increase or decrease.
- It can be measured by the Spearman Rank Order Correlation Co efficient (SROCC).
- The algorithm shows the steps to compute the SROCC.

Algorithm of SROCC

1. MOS vector x_i is sorted in descending order and the indices of the sorted vectors are assigned as rank values for the original scores. The rank vector is called X_i
2. Objective score vector y_i is sorted in descending order and the indices of the sorted vectors are assigned as rank values for the original scores. The rank vector is called Y_i
3. SROCC is given by r_s

$$r_s = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2}\sqrt{\sum(Y_i - \bar{Y})^2}}$$

Concluding Remarks

- The gist of our discussion is that there are lots of challenging issues which arises while dealing with the VQA formed synthetically.
- Primary issue is the depth while the gamer is playing.
- Furthermore, the camera moves a lot in 3D graphic games. There are lots of binocular visions.
- These core issues makes it a cutting edge research because even while assessing subjectively , each and every factor such as observer selection is very crucial.
- Also, it is hard for the state of art 2D objective metrics to yield good correlation for gaming videos.
- Evidently, we come across several artifacts in our test videos arising due to different types of distortions in rendering.

Future Work

- Seemingly, it is hard to fathom that a single quality evaluation metrics can deal with all kinds of artifacts.
- There is a need of robust objective metrics which can deal with video games precisely and predict the MOS obtained through subjective experiments dedicated especially for video games.

Thanking you for listening