Quality of Mobile Multimedia Experience

Hans-Jürgen Zepernick and Ulrich Engelke
Blekinge Institute of Technology
School of Engineering
Department of Electrical Engineering
E-mail: hans-jurgen.zepernick@bth.se
Outline

• Introduction and background
• Perception of image quality
• Perception of video quality
• QoE classifications
• Engineering approach
• Applications for mobile multimedia
• Conclusions
Introduction (I)

**Wireless Communication**

- Information transfer without using wires.
- Bandwidth is scarce, in particular, in mobile radio communication.
- Thus, signals have to be compressed prior to transmission.
- The error-prone fading channel is much more unreliable than its wired counterpart.
- Quality of Service (QoS) monitoring for end-user satisfaction.
Introduction (II)

Visual Communication

- Enabled through deployment of third generation wireless communication networks (UMTS, CDMA2000).
- High bandwidth requirements for visual content.
- Artifacts through heavy compression and error-prone channel.
- End-to-end visual quality metrics for QoS monitoring.
- To support resource management and link adaptation.
Motivation (I)

• Quality of Service (QoS) assessment is important in delivery of satisfactory quality at minimum cost

• Satisfactory mobile multimedia quality depends on who the final judge of quality is – people

• Measure and control of multimedia quality based on how human senses perceive service quality
  – Speech, audio, image, video, streaming on-demand

• Application for radio resource management and control of mobile multimedia service quality
Motivation (II)

Original

Flipped Version

PSNR=58.48dB
HIQM=24.2

Fidelity
Quality

PSNR=14.59dB
HIQM=24.2
Classification of Quality Assessment Methods

Perceptual Quality Assessment

Subjective

Objective

Psychophysical

Engineering

Intrusive

Non-intrusive

Full-reference

Reduced-reference

No-reference
Psychophysical vs Engineering Approach

- Primarily based on incorporation of various aspects of the human visual system (HVS).
- Good correlation with subjective quality perception.
- Usable over a wide range of applications.
- Generally computationally expensive.
- Data from psychophysical experiments needed.

Source: http://webvision.med.utah.edu/imaacowv/
Psychophysical vs Engineering Approach

**Engineering Approach**
- Primarily based on feature extraction.
- MOS can be used to
  - verify the impact of features on human visual perception.
  - derive mapping functions to account for non-linear quality processing.
- Less computationally expensive.
- More application specific.

![Diagram](image_url)
Example - Tiffany

• HIQM baseline: 5.00

PSNR = 23.17 dB
HIQM = 9.90 → |Δ_{HIQM}| = 4.90

PSNR = 24.15 dB
HIQM = 15.00 → |Δ_{HIQM}| = 10.00
Spatial Artifacts (I)

Blocking

Intensity masking and ringing
Spatial Artifacts (II)

Blur

Extreme artifacts
Example - Video

Distortions in salient region

Distortions in non-salient region
Temporal Artifacts

• Delay
• Frame freeze
• Frame skip
• Jerkiness
• Temporal activity
• Temporal distortions
QoE and QoS

Quality of experience

Technical factors
QoS

Non-technical factors
Subjective

QoS ← Service integrity ← QoE

Throughput, delay, jitter, data loss, bit error rate, etc.

Networking ← Middleware ← Application
Engineering Approach
Availability of Reference Information (I)

Full-reference Metrics

- Available reference image/video usually provides good quality prediction performance.
- Measure of quality degradation of test image/video as compared to the reference image/video.
- Not applicable in a communication system since reference image/video is not available at the receiver.
Availability of Reference Information (II)

**No-reference Metrics**

- Quality prediction solely based on test image/video.
- Absolute quality measure.
- Readily applicable in a communication system.
- No problem for the HVS to perform NR quality judgement but a highly difficult task for a machine.
- Cannot measure quality degradations induced during transmission.
Availability of Reference Information (III)

**Reduced-reference Metrics**

- Extracted features from reference and test image/video used for quality assessment.
- Measure of quality degradation of test image/video as compared to the reference image/video.
- Readily applicable in a communication system.
- Reduced-reference overhead becomes critical metric design issue.
Reduced-Reference Quality Assessment

Transmitter

\[ I_t \]

Feature Extraction → Pooling → Source Encoding → RR Embedding → Channel Encoding → Modulation

Receiver

Wireless Channel → De-modulation → Channel Decoding → RR Recovery → Source Decoding → Feature Extraction → Pooling

Quality metric

www.bth.se
Objective Perceptual Quality Measures

• Automatic assessment of user perception

• Artifacts
  – Blocking
  – Blur
  – Ringing
  – Masking
  – Lost block

• Example – Blur
  – Observed as texture blur
Example - Blur
Artifact Extraction

• Algorithms for extracting image artifacts

• Considered artifacts
  – Blocking artifact metric
  – Blur metric
  – Image activity metric
    • Edge activity
    • Gradient activity
      – Intensity masking detection

• Weighted sum of quality metrics
  – User perception
  – Processing complexity

$$\text{IAM}_{\text{grad}} = \frac{1}{M \cdot N} \left[ \sum_{i=1}^{M-1} \sum_{j=1}^{N} |I(i, j) - I(i+1, j)| - \sum_{i=1}^{M} \sum_{j=1}^{N-1} |I(i, j) - I(i, j+1)| \right]$$
Hybrid Image Quality Metric

- HIQM

\[ HIQM = \sum_{i=1}^{5} w_i \cdot f_i \]

where \( w_i \) is the weight of the i-th artifact and

\[ f_1 \triangleq \text{Blocking metric} \]
\[ f_2 \triangleq \text{Blur metric} \]
\[ f_3 \triangleq \text{Edge-based image activity metric} \]
\[ f_4 \triangleq \text{Gradient-based image activity metric} \]
\[ f_5 \triangleq \text{Intensity masking metric} \]
Subjective Experiments (I)

- Weights determined by subjective tests
- ITU-R Rec. BT.500-11
- Double-stimulus continuous quality-scale

Laboratory environment
- Room equipped with two 17" CRT monitors Sony CPD-E200
- Ratio of luminance of inactive screen to peak luminance was kept below a value of 0.02
- Ratio of the luminance of the screen given it displays only black level in a dark room to the luminance when displaying peak white was approximately 0.01
- Display brightness and contrast was set up via PLUGE
- Calibration of the screens was performed with the calibration equipment ColorCAL from Cambridge Research System Ltd
- DisplayMate software was used as pattern generator
- Test software for timing and presenting the test material on screen was an in-house development written in Java
Subjective Experiments (II)

- **Experiment – Australia**
  - 30 non-expert test subjects
  - 1st session only ‘Lena’ with different artifacts
  - 2nd session showing heterogenous images
  - Accounts for cross image quality

- **Experiment – Sweden**
  - 30 non-expert test subjects
  - 1st session only low texture human faces and profiles
  - 2nd session showing more complex structures
  - Accounts for impact of textures

---

**Figure 6.5:** Reference images used in Subjective Experiment 1: (a) Lena, (b) Goldhill, (c) Elaine, (d) Pepper, (e) Tiffany, and (f) Mandrill.

**Figure 6.6:** Reference images: (a) Barbara, (b) Elaine, (c) Tiffany, (d) Goldhill, (e) Pepper, and (f) Mandrill.
Subjective Experiments (III)

Presentation structure of test session

Training sequence(s)

Stabilizing sequence(s) (results for these items are not processed)

Break (to allow time to answer questions from observers)

Main part of test session

Source: ITU-R BT.500-12
Subjective Experiments (IV)

Presentation structure of test material

Phases of presentation:

T1 = 10 s  Test sequence A
T2 = 3 s   Mid-grey produced by a video level of around 200 mV
T3 = 10 s  Test sequence B
T4 = 5-11 s Mid-grey

Source: ITU-R BT.500-12
Subjective Experiments (V)

Portion of quality-rating form using continuous scales

<table>
<thead>
<tr>
<th></th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Excellent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Good</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bad</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: ITU-R BT.500-12
Subjective Experiments (VI)

ITU-R quality and impairment scales

<table>
<thead>
<tr>
<th>Quality</th>
<th>Impairment</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Excellent</td>
</tr>
<tr>
<td>4</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td>Fair</td>
</tr>
<tr>
<td>2</td>
<td>Poor</td>
</tr>
<tr>
<td>1</td>
<td>Bad</td>
</tr>
<tr>
<td>5</td>
<td>Imperceptible</td>
</tr>
<tr>
<td>4</td>
<td>Perceptible, but not annoying</td>
</tr>
<tr>
<td>3</td>
<td>Slightly annoying</td>
</tr>
<tr>
<td>2</td>
<td>Annoyng</td>
</tr>
<tr>
<td>1</td>
<td>Very annoying</td>
</tr>
</tbody>
</table>

Source: ITU-R BT.500-12
Subjective Experiments (VII)

DEMO
Evaluation (I)

- Ratings averaged into mean opinion score (MOS)
- Prediction curve
- Mapping
Evaluation (II)

- Pearson linear correlation coefficient

\[ r_p = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \]

where

\( x_i, y_i \): Predicted rating and the subjective rating

\( \bar{x}, \bar{y} \): Means of the respective data sets

- Spearman rank order correlation coefficient

\[ r_s = \frac{\sum_i (P_i - \bar{P})(S_i - \bar{S})}{\sqrt{\sum_i (P_i - \bar{P})^2} \sqrt{\sum_i (S_i - \bar{S})^2}} \]

where

\( P_i, S_i \): Ranks of the predicted scores and the subjective scores

\( \bar{P}, \bar{S} \): Midranks of the respective data sets

- Outlier ratio

\[ R_O = N_O / N \]

where \( N_O \) is the total number of outliers and \( N \) is the data set size

- Exponential prediction function to predict MOS of metric \( \mu \)

\[ MOS_{\mu} = ae^{b\mu} \]
Evaluation (III)

• Subjective Experiment 1

<table>
<thead>
<tr>
<th>Feature/Artifact</th>
<th>Metric</th>
<th>Weight</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocking</td>
<td>$\Delta_{\text{Block}}$</td>
<td>$w_1$</td>
<td>0.81</td>
</tr>
<tr>
<td>Blur</td>
<td>$\Delta_{\text{Blur}}$</td>
<td>$w_2$</td>
<td>0.56</td>
</tr>
<tr>
<td>Edge-based activity</td>
<td>$\Delta_{\text{LAM}_{\text{Edge}}}$</td>
<td>$w_3$</td>
<td>0.77</td>
</tr>
<tr>
<td>Gradient-based activity</td>
<td>$\Delta_{\text{LAM}_{\text{Grad}}}$</td>
<td>$w_4$</td>
<td>0.20</td>
</tr>
<tr>
<td>Intensity masking</td>
<td>$\Delta_{\text{Mask}_{\text{Int}}}$</td>
<td>$w_5$</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Prediction function: $f(x) = 77e^{-0.203x}$, $x = \Delta_{\text{HIQM}}$

• Subjective Experiment 2

<table>
<thead>
<tr>
<th>Feature/Artifact</th>
<th>Metric</th>
<th>Weight</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocking</td>
<td>$\Delta_{\text{Block}}$</td>
<td>$w_1$</td>
<td>0.77</td>
</tr>
<tr>
<td>Blur</td>
<td>$\Delta_{\text{Blur}}$</td>
<td>$w_2$</td>
<td>0.35</td>
</tr>
<tr>
<td>Edge-based activity</td>
<td>$\Delta_{\text{LAM}_{\text{Edge}}}$</td>
<td>$w_3$</td>
<td>0.61</td>
</tr>
<tr>
<td>Gradient-based activity</td>
<td>$\Delta_{\text{LAM}_{\text{Grad}}}$</td>
<td>$w_4$</td>
<td>0.16</td>
</tr>
<tr>
<td>Intensity masking</td>
<td>$\Delta_{\text{Mask}_{\text{Int}}}$</td>
<td>$w_5$</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Prediction function: $f(x) = 96.15e^{-0.2075x}$, $x = \Delta_{\text{HIQM}}$

• Combined Results

<table>
<thead>
<tr>
<th>Feature/Artifact</th>
<th>Metric</th>
<th>Weight</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocking</td>
<td>$\Delta_{\text{Block}}$</td>
<td>$w_1$</td>
<td>0.79</td>
</tr>
<tr>
<td>Blur</td>
<td>$\Delta_{\text{Blur}}$</td>
<td>$w_2$</td>
<td>0.45</td>
</tr>
<tr>
<td>Edge-based activity</td>
<td>$\Delta_{\text{LAM}_{\text{Edge}}}$</td>
<td>$w_3$</td>
<td>0.70</td>
</tr>
<tr>
<td>Gradient-based activity</td>
<td>$\Delta_{\text{LAM}_{\text{Grad}}}$</td>
<td>$w_4$</td>
<td>0.18</td>
</tr>
<tr>
<td>Intensity masking</td>
<td>$\Delta_{\text{Mask}_{\text{Int}}}$</td>
<td>$w_5$</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Prediction function: $f(x) = 84.2e^{-0.2327x}$, $x = \Delta_{\text{HIQM}}$
Objective Metric Design Framework
Non-Linear Regression

**Non-linear Visual Quality Processing**

- Mapping of metric values onto predicted MOS $MOS_p$.
- Distinction between quality levels is more pronounced in regions of small degradations $\iff \Delta_{D1} = \Delta_{D2}$ we have $\Delta_{Q1} > \Delta_{Q2}$.
Multi-Resolution Structural Degradation Metrics for Perceptual Image Quality Assessment

**Background**

- Visual processing at multiple scales in the HVS.
- Artifacts may be perceived differently at certain resolutions.
- Multi-resolution analysis of artifacts incorporated into quality metric design.
Multi-Resolution Analysis (I)

**Gaussian Pyramid Decomposition**

- Gaussian pyramid mirrors the multiple scale processing in the HVS.
- Iterative Gaussian filtering:

\[
g_l(x, y) = \sum_{u=-2}^{2} \sum_{v=-2}^{2} \sigma(u, v) \cdot g_{l-1}(2x + u, 2y + v)
\]
Multi-Resolution Analysis (II)

**Visualisation of Pyramid Decomposition**

- Level $g_0$
- Level $g_1$
- Level $g_2$
- Level $g_3$
- Level $g_4$
- Level $g_5$
Quality Assessment Using Multi-Scale Features

Transmitter

Receiver

Source Encoding

RR Embedding

Wireless Link

Source Decoding

RR Recovery

Gaussian Pyramid

Feature Extraction $g_0$

Feature Extraction $g_1$

...$

Feature Extraction $g_{L-1}$

Multiple-scale Metric Computation

$Q_{MS}$

$I_r$

$I_d$
Numerical Results

**Prediction Accuracy for G2**

- Prediction accuracy tends to increase with the number of levels.
- Perceptual feature relevance weights, $w_p$, further improve accuracy.
- Strongest increase up to level $g_3$.
- Degradation information for each level is incorporated in the metric.
Saliency Awareness for Objective Video Quality Assessment (I)

Original video frame

Source: Engelke, Barkowsky, Le Callet, Zepernick, QoMEX 2010
Saliency Awareness for Objective Video Quality Assessment (II)

Scatter plot of TetraVQM

Prediction performance

Source: Engelke, Barkowsky, Le Callet, Zepernick, QoMEX 2010
Mobile Multimedia Applications

• Scope
  – Interactive mobile multimedia
  – Mobile video streaming on-demand
  – Mobile television

• Challenges
  – Complex traffic patterns
  – Heterogeneous network structures
  – Severe channel impairments

• Aim
  – Perceptual-based real-time quality assessment
  – Radio resource management and quality control
Applications (I)

• Unequal Error Protection over Fading Channel
Applications (II)

• De-blocking filter for H.263 mobile video sequences

Source: Engelke, Rossholm, Zepernick, Lövström, ISWPC 2007
In-Service Quality Monitoring

- Beneficial in generating video frames with a wide range of artifacts
- Rayleigh fading in the presence of AWGN
- $\frac{E_b}{N_0}=5$ dB
- BPSK and $(31,21)$ BCH code
- H-ARQ with soft-combining
- Maximum number of retransmissions is set to 5
- Total of 100 consecutive frames of uncompressed quarter common intermediate (QCIF) videos
- Compressed at bit rate of 1bpp using Kakadu software
- MJ2 video were sent over channel and decompressed to QCIF at receiver
Frame samples of the video “Highway drive” (I)

- Artifacts

Intensity masking

Blocking

Frame no. 2

Frame no. 33

Frame no. 80

Frame no. 89
Frame samples of the video “Highway drive” (II)

• Progression of the metrics
Some Characteristics

- HIQM and SSIM provide distinctive quality metrics
- HIQM (RR) identifies the same frames as of perceptually lower quality as those detected by SSIM (FR)
- HIQM accounts better for blocking artifacts as SSIM
- RRIQM is rather unstable in providing quality level

**Overhead and Computational Complexity (100 frames of MJ2 video “Highway drive”)**

<table>
<thead>
<tr>
<th></th>
<th>$\Delta_{HIQM}$</th>
<th>RRIQA</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overhead Processing</td>
<td>0.07%</td>
<td>0.65%</td>
<td>100%</td>
</tr>
<tr>
<td>time (sec)</td>
<td>51.27</td>
<td>380.39</td>
<td>3.75</td>
</tr>
</tbody>
</table>
Conclusions

• Psychophysical versus engineering approach for mobile multimedia quality assessment

• Illustration of mechanisms that can impair viewing experience

• Classification of QoE and relationship to QoS

• Artifact extraction for images and video

• Applications in mobile multimedia services
  – Improved quality of mobile multimedia
  – Conservation of radio resources
Ongoing and Future Activities (I)

- Temporal artifacts
- Algorithms for extracting temporal artifacts
- Algorithms for dealing with lost blocks
- Subjective experiments for wireless video
- Subjective experiments for relationship between audio and video quality perception
- Evaluation of approach on real data
- Application to region of interest
- Region of interest and cooperative communications
- Perceptual-based cross-layer design of mobile video systems
Ongoing and Future Activities (II)

DEMO