Correlating Video Quality Metrics to User Experience: an Event-based Approach

- Master’s Thesis Defense -

Yongfeng Huang

Supervisor: Prof. James Won-Ki Hong

Division of ITCE, POSTECH, Korea

xgjonathan@postech.ac.kr

2012.06.22
Contents

- Introduction & Motivation
- Background & Related Work
- Event-based user experience assessment
  - Classification of defect events
  - Event to experience correlation
- Validation
- Conclusion & Future Work
Internet Traffic Trend

- Multimedia will dominate IP traffic

The Management Challenge

❖ Management of multimedia services
  - For content providers, ensuring service quality is important (*service differentiator to attract and retain customers*)
  - To be specific, network service providers need to assess, manage, and improve user perceptual experience

❖ User experience of video quality
  - Compared to quality of service (QoS), quality of experience (QoE) considers human perception
  - Human vision system (HVS) is very complicated to model, i.e. user’s preference for video content, user’s interest of frame region, etc.
Current Limitations

- **User feedback is impractical in reality**
  - Well-controlled user experiment with a wide variety of test elements is not easy to design or administer
  - Users cannot be bothered with giving truthful feedback each time

- **Limitation of current objective measurements**
  - Temporal factors are missed, i.e. the peak signal-to-noise ratio (PSNR)
  - Content-related factors are missed
  - Events play a crucial role in human experience (*not reflected*) [J. M. Zacks et al., 2010]
Objectives

Propose an approach for correlating video quality metrics to user experience
- Accurate and efficient

Consider event-based human experience
- How to identify/segment defect events in videos
- How to classify defect events
- How to correlate events to user experience
Contributions

- An event-based approach for correlating video quality metrics to user experience
  - The first (to the best of my knowledge) to consider that human experience is event-based
  - Event classification: accurate and efficient classification of defect events (eSSIM Aggregator)
  - Event-to-experience correlations: correlate different event types to user experience by constructing event-specific user models.
Background

- **VIDAR**: video quality analyzer in real-time
  - [A. Kwon et al., 2012]
  - Correlate network performance to user experience
  - Structure Similarity (SSIM) – frame quality metric
    - [Z. Wang et al., 2004]

Diagram:

- **Network**
  - QoS
  - Loss
  - Delay
  - Error
  - Jitter
  - Error Correlation
  - R3 Analysis
  - eSSIM of frames

- **Service**
  - Subjective Filters
  - Luma Adjust
  - Frame Complexity
  - Scene Change
  - eSSIM Aggregator (Machine Learning)

- **Customer**
  - User Model
  - Estimated MOS
  - Vidi of events
  - User Model

**R3 model**

**Vidi model**

**Machine Learning (ML)**

**Mean opinion score (MOS)**
Related Work

*Statistical approach*
- Rely on one or two salient perception characteristics
- Mean square error (MSE) [A. Bhat et al., 2009]
- PSNR [O. Nemethova et al., 2006]
- Content factors are missed

*Machine learning approach*
- Decision tree [V. Menkovski et al., 2009]
  - Data instances: video spatial information, video temporal information, frame rate and bit rate
  - Result: “YES” or “NO”
Contents

- Introduction & Motivation
- Background & Related Work
- Event-based user experience assessment
  - Classification of defect events
  - Event to experience correlation
- Validation
- Conclusion & Future Work
Process Flow of eSSIM Aggregator

- eSSIM aggregator for detecting and classifying defect events

**Input:** eSSIM of frames

**Output:** defect events
Preprocessing (Discontinuity Filter)

- Meaning of eSSIM values
- Show the intensity of discontinuity in the video

Foreman, GOP = 12, 2% GE loss model
Defect Event Segmentation

❖ A set of rules to segment defect events
  • i.e. the first 10 frames of each video are ignored

Foreman, GOP = 12, 2% GE loss model
Feature Extraction and Normalization

❖ Purposes of extracting features
  - To distinguish different types of defect events
  - Reduce the dimension and input of ML classifiers

❖ Features extracted
  - Mean ($\mu$) and standard deviation ($\sigma$)
  - Minimum: the minimum eSSIM value of an event
  - Defective ratio:
    \[ \text{ratio} = \frac{N_{\text{essim} < 0.95}}{n} \]
  - Severity of dropped and duplicated frames:
    \[ \text{severity} = \frac{N_{\text{essim} \leq 0.0}}{n} \]
  - Skewness: measure of the asymmetric of the probability distribution of event data
  - Kurtosis: measure of whether the distribution is peaked or flat, relative to a normal distribution

❖ Feature normalization
  - Middle-range normalization: [-1, 1]
  - Mean-std normalization: mean = 0, std = 1
Multi-class Classifiers - Types of Defect Events

- **Distortion**: a series of frames containing perceivable distortions

- **Glitch**: a series of frames, where distortion is short and slightly perceivable

- **Freezing**: a series of duplicated frames

- **Video samples of defect events**
Multi-class Classifiers

- **Support vector machine (SVM)**
  - Multi-dimension and continuous features
  - Nonlinear situations

- **Multi-class classification with SVM**
  - Combination of binary SVM classifiers
    - One-versus-all:
    - One-versus-one:
Correlating event metrics to User Experience

- Correlational models for predicting user MOS from event metrics

- Intensity features for Distortion
  \{\mu, \text{ratio}, \text{frame complexity, motion speed}\}

- Intensity features for Glitch
  \{\mu, \text{ratio}, \text{frame complexity, motion speed}\}

- Intensity features for Freezing
  \{\text{the time duration, discontinuity intensity}\}

- Artificial neural network (ANN)
  - No readily discernable patterns statistically
  - ANN shows good accuracy
Contents

❖ Introduction & Motivation
❖ Background & Related Work
❖ Event-based user experience assessment
  ◦ Classification of defect events
  ◦ Event to experience correlation
❖ Validation
❖ Conclusion & Future Work
Validation

> **Experiment setup**

<table>
<thead>
<tr>
<th>Name</th>
<th>Number of Scene Cuts</th>
<th>Scene Moving Speed</th>
<th>Object Moving Speed</th>
<th>Average Frame Complexity</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bus</td>
<td>0</td>
<td>Middle</td>
<td>Fast</td>
<td>27.95</td>
<td>Bus</td>
</tr>
<tr>
<td>Container</td>
<td>0</td>
<td>No</td>
<td>Slow</td>
<td>16.18</td>
<td>Ship with containers</td>
</tr>
<tr>
<td>Flower</td>
<td>0</td>
<td>Middle</td>
<td>Slow</td>
<td>22.14</td>
<td>Flower and house</td>
</tr>
<tr>
<td>Football</td>
<td>0</td>
<td>Fast</td>
<td>Fast</td>
<td>11.96</td>
<td>Football players</td>
</tr>
<tr>
<td>Foreman</td>
<td>0</td>
<td>Slow</td>
<td>Slow</td>
<td>18.73</td>
<td>Portrait and construction</td>
</tr>
<tr>
<td>Mother &amp;Daughter</td>
<td>0</td>
<td>No</td>
<td>Slow</td>
<td>12.60</td>
<td>Portrait</td>
</tr>
<tr>
<td>Stefan</td>
<td>0</td>
<td>Slow</td>
<td>Middle</td>
<td>23.62</td>
<td>Tennis player</td>
</tr>
<tr>
<td>Inception</td>
<td>6</td>
<td>Fast</td>
<td>Fast</td>
<td>18.00</td>
<td>Portrait, construction and etc.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distortion</td>
<td>151</td>
</tr>
<tr>
<td>Glitch</td>
<td>39</td>
</tr>
<tr>
<td>Freezing</td>
<td>48</td>
</tr>
<tr>
<td>Total</td>
<td>238</td>
</tr>
</tbody>
</table>
Kernel Selection and Parameter Setting (SVM)

- **Sequential minimal optimization (SMO)** (J. Platt et al., 1998)
  - An algorithm for efficiently solving the optimization problem that arises during the training of SVMs

- **Kernel selection**
  - Linear kernel
  - Radial basis function (RBF) kernel

- **Parameter setting (RBF)**
  - Grid search
  - Optimized method (S. Keerthi et al., 2003)

1. **SMO-L**: SMO with **Linear** kernel
2. **SMO-G**: SMO with RBF kernel, and **Grid** search for parameters
3. **SMO-O**: SMO with RBF kernel, and **Optimized** method for parameters
Efficiency of multiclass classification
### Classification result of each binary classifier

<table>
<thead>
<tr>
<th>Distortion vs. Glitch</th>
<th>Distortion vs. Freezing</th>
<th>Glitch vs. Freezing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>92.04%±0.24%</td>
<td>85.38%±0.32%</td>
</tr>
</tbody>
</table>

**Diagrams:**

- **Foreman:** GOP = 12, 2% GE loss model, Frame complexity = 18.73
- **Flower:** GOP = 12, 2% uniform loss model, Frame complexity = 22.14
### Subjective testing of user model

<table>
<thead>
<tr>
<th>MOS</th>
<th>1, 2</th>
<th>3, 4</th>
<th>5, 6</th>
<th>7, 8</th>
<th>9, 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality</td>
<td>Bad</td>
<td>Poor</td>
<td>Fair</td>
<td>Good</td>
<td>Excellent</td>
</tr>
<tr>
<td>Impairment</td>
<td>Very annoying</td>
<td>Annoying</td>
<td>Slightly annoying</td>
<td>Perceptible but not annoying</td>
<td>Imperceptible</td>
</tr>
</tbody>
</table>

### Average user MOS of defect events

MOS range of Distortion: 1, 2, 3, 4, 5
MOS range of Glitch: 6, 7, 8, 9
MOS range of Freezing: 3, 4, 5, 6, 7
## Classification result of user MOS on Distortion

<table>
<thead>
<tr>
<th>User MOS</th>
<th>Classified MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 10 4 0 0</td>
</tr>
<tr>
<td>2</td>
<td>11 11 0 1</td>
</tr>
<tr>
<td>3</td>
<td>2 13 21 2 1</td>
</tr>
<tr>
<td>4</td>
<td>0 1 5 7 0</td>
</tr>
<tr>
<td>5</td>
<td>0 3 11 1 0</td>
</tr>
</tbody>
</table>

(a) \{\mu, \text{ratio}\}

<table>
<thead>
<tr>
<th>User MOS</th>
<th>Classified MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10 2 0 0</td>
</tr>
<tr>
<td>2</td>
<td>17 11 1 1</td>
</tr>
<tr>
<td>3</td>
<td>11 19 6 2</td>
</tr>
<tr>
<td>4</td>
<td>2 6 4 1</td>
</tr>
<tr>
<td>5</td>
<td>3 8 4 0</td>
</tr>
</tbody>
</table>

(b) \{\mu, \text{ratio, frame complexity}\}

<table>
<thead>
<tr>
<th>User MOS</th>
<th>Classified MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3 8 4 0 0</td>
</tr>
<tr>
<td>2</td>
<td>17 11 0 1</td>
</tr>
<tr>
<td>3</td>
<td>10 21 5 3</td>
</tr>
<tr>
<td>4</td>
<td>2 4 5 2</td>
</tr>
<tr>
<td>5</td>
<td>2 8 3 2</td>
</tr>
</tbody>
</table>

(c) \{\mu, \text{ratio, frame complexity, motion speed}\}

<table>
<thead>
<tr>
<th></th>
<th>Exact</th>
<th>Drift ±1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>34%</td>
<td>79%</td>
</tr>
<tr>
<td>(b)</td>
<td>37%</td>
<td>83%</td>
</tr>
<tr>
<td>(c)</td>
<td>41%</td>
<td>83%</td>
</tr>
</tbody>
</table>
Conclusion

❖ **Summary**
  - An event-based model to correlate video quality metrics to user experience, using ML
  - Event classification and event-to-experience correlations

❖ **Contribution**
  - The **first** (to the best of my knowledge) to consider that human experience is event-based
  - Event classification (eSSIM aggregator)
    • Provides an accurate classification of defect events with low training time
  - Relating to user experience
    • Shows a strong correlation between key features (specific to each event type) and user MOS
  - A good method for problem decomposition
Future Work

- **User MOS estimation for entire video session**
  - Current user MOS estimation is at the event level
  - A number of psychological theories will be used for modeling the complicated human perception system

- **Reference model of user perception**
  - Find common response among users to generate reference model
  - Identify key user-specific features that specialize reference model to individual user (or groups of like-minded users) MOS
Thank You!
Reference


Related Work – Machine Learning Techniques

 Performance of ML
  • **Accuracy**: A strong correlation between the input parameters (i.e., features) and the output result
  • **Efficiency**: The number of features should be small and indicative

 Comparison of supervised ML algorithms
  • Logic-based algorithms, i.e. decision tree
    • Do not perform well with numerical features
  • Perceptron-based algorithms, i.e. artificial neural network
    • Inefficient with the presence of irrelevant features
  • Statistical learning algorithms, i.e. k-nearest neighbor (k-NN)
    • Very sensitive to irrelevant features
  • Support vector machine (SVM)
    • Performs well with multi-dimension and continuous features, as well as when a nonlinear relationship exists among the input and output features
Event segmentation

- **Rationales:**
  - Human experience is event-based
  - The length of event can be scalable and the boundary between events can be coarse

- **Rules:**
  - The minimum length of a defect event is 10
  - The maximum length of a defective event is 100, about 3 seconds
  - The minimum length of boundary between two events is 10
  - When $eSSIM \geq 0.95$, distortion is too small to be perceived. Therefore, we regard these frames as normal ones with $eSSIM = 1$
  - The first 10 frames of each video are ignored, because they can be counted as frames after a scene change
Background (2/2)

- **H.264/AVC**
  - Use spatial and temporal redundancy for compression

- **Measuring user experience**
  - Mean opinion score (MOS)

- **Machine learning (ML) techniques**
  - ML has been widely used in literature to establish correlation among metrics that are:
    - complex to mathematical model or statistically correlate
    - Containing many inter-dependent parameters (difficult to isolate and deduce from ground data)
**Feature sensitivity**

Dashed line: accuracy without feature excluded

Solid line: accuracy with feature excluded

Glitch vs. Freezing

Glitch vs. Distortion

Distortion vs. Freezing

Accuracy vs. Feature Excluded:
- skewness
- severity
- δ
- minimum
- μ
- kurtosis
- ratio
Preprocessing for eSSIM raw data

- Content filters (A. Kwon et al., 2012): luminance, frame complexity, scene change and motion

- **Discontinuity filter**
  - Frames dropped
  - Frames duplicated

\[
eSSIM_{\text{disc}} = \text{ssim} (\text{frame}_b, \text{frame}_a) - 1,
\]

- \(eSSIM_{\text{disc}}\) has a range of \([-1, 0]\)

- Points need to mention
  - The process of discontinuity filter is done at the server side
  - Scene change remedy: the first few frames following a scene change are not perceived (A. R. Reibman et al., 2007)
Classification result of user MOS on Glitch and Freezing

<table>
<thead>
<tr>
<th>User MOS</th>
<th>Classified MOS</th>
<th>Classified MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

(a) Glitch
16/26, 23/26

(b) Freezing
16/31, 30/31
User Experiment

- A single stimulus procedure
- A random playlist
eSSIM Aggregator (9/)

- **Features extracted**
  - $\mu$ and $\delta$:
  - Minimum: the minimum eSSIM value of an event
  - Defective ratio:
    \[ \text{ratio} = \frac{N_{eSSIM<0.95}}{n} \]
  - Severity of dropped and duplicated frames:
    \[ \text{severity} = \frac{N_{eSSIM\leq0.9}}{n} \]
  - Skewness:
    - Measure of the asymmetric of the probability distribution of event data
    - $[-1.0, 0.0, 0.1, 0.5, 0.8, 0.9, 0.95, 0.98, 1.0]$ 
  - Kurtosis: measure of whether the data are peaked or flat, relative to a normal distribution

- **Preprocessing for features**
  - Middle-range normalization: $[-1, 1]$
  - Mean-std normalization: mean = 0, std = 1
Objective metrics

- **SSIM**
  - Full reference metric
  - Human visual perception is highly adapted for extracting structural information from a scene
  - Structural information is the idea that the pixels have strong inter-dependencies especially when they are spatially close

- **PSNR**
  - Computed by averaging the squared intensity differences of distorted and reference image pixels
  - Inconsistent with human eye perception
Statistical approach

A. Bhat et al., 2009

\[ MOS_p = 1 - k(MSE), \]

• Where MSE is the mean squared error, and k is derived from the spatial edge strength.
• Do not consider distortion caused by varying network conditions

O. Nemethova et al., 2006

\[ \overline{MOS}_{PSNR}[n] = a \cdot PSNR[n] + b, \]

• Where n is frame index, a and b are derived from the relationship between PSNR and MOS
• Frame-level metric, full-reference
eSSIM Aggregator (10/)

- **Multiclass classification with SVM**
  - Kernel selection
    - Radial basis function (RBF)
  - Combination of binary SVM classifiers
    - One-versus-all: $M$
    - One-versus-one: $M(M-1)/2$
  - Parameters setting:
    - $C$ and $\gamma$

\[ k(\hat{x}, \bar{x}_i) = \exp(-\gamma \| \hat{x} - \bar{x}_i \|^2), \]

A. Ben-Hur, “A user’s guide to support vector machines,”