VIDAR: Video Quality Analyzer in Real-Time
- Master’s Thesis Defense -

Arum Kwon
Supervisor: Prof. James Won-Ki Hong
Dept. of CSE, POSTECH, Korea
arumk@postech.ac.kr
2011. 09. 30
1. Introduction & Motivation
2. Background & Related Work
3. VIDAR Approach
4. Validation
5. Conclusion & Future Work
Multimedia is becoming more and more popular

“By 2013 video will be 90 percent of all consumer IP traffic and 64 percent of mobile”

– Cisco Systems*

* Cisco Visual Networking Index: Forecast and Methodology, 2010-2015
Management of multimedia services

- Multimedia services require high network performance
- Quality of multimedia service should be measured and managed based on quality
- Popularity of mobile devices impose energy and communication constraints

Customer perceived quality

- Fundamental indicator of customer satisfaction & retention rate in multimedia services
- Hard to measure directly from network performance, because customer perception is subjective
Motivation

- **Necessity of a real-time video quality analyzer**
  - Automated process allows the operator to analyze a large set of data
  - Real-time analysis enables real-time service quality management and troubleshooting

- **Key challenges**
  - Full-reference is not feasible in real-time system
  - Image quality metrics do not match well with video quality
  - “Black-box” approach needs extensive user participation

- **Importance of network performance**
  - Major cause of video quality defects
  - A key aspect of management and planning for service operators
Research Objectives

- **Real-time video quality analyzer**
  - Compute objective metric for assessing impact of network performance on image quality
  - Identify key aspects of video quality management and their relationship
  - Compute subjective perceived video quality
  - A feasible system design
  - Validation through experiments
VIDAR (VIDeo quality Analyzer in Real-time)

- Real-time estimation of SSIM*
  - Good estimation, minimal client processing and communication overhead

Vidi, a new subjective video quality metric

- Incorporates video content characteristic
- Considers non-linear subjectivity of user evaluation
- Automatic event-based categorization and indication of perceived video defects

Validation with standard reference video and video clip from film

*SSIM: Structural Similarity (page 8)
Background

- **PSNR (Peak-Signal-to-Ratio)**
  - Easy to compute and have good reduced-reference models
  - Doesn’t match well with perceived quality

- **SSIM (Structural Similarity)**
  - Captures subjective quality than PSNR
  - Full-reference metric

- **H.264/AVC**
  - Most commonly used codec for network-based video services
  - Use spatial / temporal redundancy for compression
Related Work

- **Analytical approach (S. Tao et al., 2008)**
  - Estimated PSNR by modeling impact of packet loss to video frame defects
  - Don’t consider content and codec

- **Black-box approach (M. Venkataraman et al., 2011)**
  - Constructed a QoE (Quality of Experience) space by an extensive user experiment under different network and video parameters
  - Not feasible for a large parameter set
Overview of VIDAR

- VIDAR
  - VIDeo quality Analyzer in Real-time
  - Measure subjective video quality by relating different layers

R3 Model  Vidi Model
VIDAR - R3 Model

- **R3 model**
  - Reversed Reduced Reference model
  - Maps network performance (packet loss) to objective video quality

- **eSSIM (estimated SSIM)**
  - Estimates SSIM based on the error information
  - Exploit SSIM relations between distorted and reference frames in source video

- **eSSIM of a frame**
  - $eSSIM = 1 – \text{reconstruction error} – \text{propagation error}$
Error Propagation

- Example of frame defects caused by lost slices and error propagation

![Diagram showing example of frame defects caused by lost slices and error propagation.](image-url)
Performance of eSSIM

- eSSIM shows good correlation with SSIM

![Graph showing correlation between eSSIM and SSIM with Bernoulli Loss and GE Loss](image)
Comparing SSIM, eSSIM and PSNR

\[ \frac{Y}{\mu_y} \]

- SSIM
- eSSIM
- PSNR

Frame ID

[ SSIM, eSSIM, and PSNR under GE Loss ]
Vidi Model

Vidi model

- Adjusts eSSIM values by subjective filters
- Produces Vidi (VIDeo Index) metric to capture instances of perceived video defects

Subjective filters of Vidi model

- Luminance adjust
  - Distortions in a dark frame is less visible
- Frame complexity
  - Distortions in a complex frame is less visible
- Scene change
  - Distortions in frames immediately after scene changes are not perceived
Vidi Metric

- **Vidi metric**
  - Represent a defect event in terms of types of defects and intensity
  - Shows defective events that viewers perceive

- **Types of defects**
  - Distortion, freezing, discontinuity, stutter, and glitch

- **Intensity**
  - Complement of MOS mapped from eSSIM by a logistic regression function

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Intensity (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distortion</td>
<td>3</td>
<td>47.3</td>
</tr>
<tr>
<td>Freezing</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Discontinuity</td>
<td>1</td>
<td>57.0</td>
</tr>
<tr>
<td>Stutter</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Glitch</td>
<td>1</td>
<td>72.0</td>
</tr>
</tbody>
</table>

[Summary of Vidi Metrics]
Validation

❖ Experiment process

Server

- Source Video
- Encoder
- Encoded Video
- Packetizer
- Video Packets
- eSSIM Aggregator
- R3 Analyzer

Loss-emulated Network

Client

- Video Packets with Loss
- Modified Decoder
- Error-Correlation Info.
- Decoded Video

❖ Impact of packet loss on eSSIM

[ Bernoulli Loss ]

[ GE Loss ]
Loss rate and GOP size

- Higher loss rate causes more distortions, but the increase is not linear.
- Larger GOP size causes prolonged distortions and more frame drops, also non-linear.

Experiment Result (1/4)

[ Bernoulli Loss (2%, 4%) ]

[ 2% Bernoulli Loss (GOP 12, 24) ]
When packet loss is bursty

- Bursty loss causes many frame discards in short time period (i.e., defects are concentrated)
- Duration is shorter with frequent scene changes (due to shorter GOP sizes)

Experiment Result (2/4)
Video Example

(a)

(b)
Experiment Result (3/4)

[ (a) Foreman under Bernoulli Loss ]

Surveyed MOS: 3

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Intensity (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distortion</td>
<td>3</td>
<td>57.0</td>
</tr>
<tr>
<td>Freezing</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Discontinuity</td>
<td>1</td>
<td>57.0</td>
</tr>
<tr>
<td>Stutter</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Glitch</td>
<td>4</td>
<td>69.0</td>
</tr>
</tbody>
</table>

[ Summary of Vidi Metrics in (a) ]

[ (b) Inception under Bernoulli Loss ]

Surveyed MOS: 4

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Intensity (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distortion</td>
<td>2</td>
<td>59.0</td>
</tr>
<tr>
<td>Freezing</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Discontinuity</td>
<td>14</td>
<td>22.0</td>
</tr>
<tr>
<td>Stutter</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>Glitch</td>
<td>2</td>
<td>75.0</td>
</tr>
</tbody>
</table>

[ Summary of Vidi Metrics in (b) ]
Impact of packet loss

- Packet loss is a major cause of video quality degradation
- Packet loss rate and loss pattern (uniform or bursty) are important parameters and their effects are non-linear

Impact of GOP size

- GOP size limits propagation error, the actual GOP size is highly content dependent (scene change)

Effectiveness of Vidi metric

- Vidi metric captures instances of perceived video defects by type and intensity
- Vidi metric estimate MOS much better than image quality metrics, such as PSNR and SSIM
  - Many cases where SSIM patterns do not match MOS
Summary

- **VIDAR, a new approach for assessing perceived video quality**
  - Examined the correlations among network performance, objective video quality, and perceived video quality

- **R3 model**
  - eSSIM can well estimate SSIM in an effective and efficient way

- **Vidi model**
  - Produces subjective video defects measured by subjective filters
  - Vidi metric represents perceived video quality effectively, and shows better relation with MOS
Future Work

- **User model**
  - User experience is a process of recollecting the past reactions to events
  - User experience can be measured using Vidi metrics of a video
  - Model for showing the process of recollecting reactions is needed

- **More validation**
Q&A

THANK YOU


Classification of Video Quality Assessment

- Full-reference
- Reduced-reference
- No-reference
Group Of Pictures

Group Of Pictures (GOP)

- Coded video frames are grouped into GOPs
- I-frame (Intra frame)
- P-frame (Predictive frame)
- B-frame (Bi-predictive frame)
H.264/AVC
Computation of eSSIM of a Frame

- **I-frame**

\[
eSSIM(I, I') = 1 - Q_3 \times \frac{\sum_{ms}(1 - SSIM_{ms}(I, I_r))}{eSSIM(I_r, I'_r) \times S_n}
\]

- **P-frame**

\[
eSSIM(P, P') = 1 - Q_2 \times \frac{\sum_{ms}(1 - SSIM_{ms}(P, P_r))}{eSSIM(P_r, P'_r) \times S_n} - \frac{1 - eSSIM(P_r, P'_r)}{[SSIM(P, P_r)]} \times \frac{\Phi}{S_n}
\]

- **B-frame**

\[
eSSIM(B, B') = 1 - Q_2 \times \left(\frac{\sum_{ms}(1 - SSIM_{ms}(B, B_{r1}))}{eSSIM(B_{r1}, B'_{r1}) \times 2S_n} + \frac{\sum_{ms}(1 - SSIM_{ms}(B, B_{r2}))}{eSSIM(B_{r2}, B'_{r2}) \times 2S_n}\right) - \frac{\Phi((1 - eSSIM(B_{r1}, B'_{r1})) + (1 - eSSIM(B_{r2}, B'_{r2})))}{2S_n}
\]

- **B-frame (one reference is missing)**

\[
eSSIM(B, B') = 1 - Q_2 \times \frac{\sum_{ms}(1 - SSIM_{ms}(B, B_{r1}))}{eSSIM(B_{r1}, B'_{r1}) \times S_n} - \frac{1 - eSSIM(B_{r1}, B'_{r1})}{SSIM(B, B_{r1})} \times \frac{\Phi}{S_n}
\]
Error-correlation Tree

- **Error-correlation tree**
  - Represent the lost slices information and the reference relationships of a GOP (Group Of Pictures)

- **Impact of a lost slice to other slices**
  - I slice: slices with a same slice number in the entire frames in the GOP
  - P slice: slices who refer the lost slice
  - B slice: no impact

```plaintext
F62 I slc_frm_oth SLC2 {SL62:1 SL62:3}
F63 P {F62} slc_frm_ref SL1
F64 B {F63 F66} slc_ok
F65 B {F63 F66} slc_frm_ref SL2
F66 P {F63} slc_drop
F67 B {F66 F69} slc_ok
F68 B {F66 F69} slc_ok
F69 P {F66} slc_ok
```
Mapping SSIM to MOS

![Graph showing the relationship between MOS and MSSIM for JPEG and JPEG2000 images. The graph includes a line of best fit using a Logistic Function.](image)
Gilbert-Elliott Model

- Gilbert-Elliott (GE) model
  - A channel model widely used for describing burst error patterns in transmission channel
  - State 0: no loss, state 1: loss
  - Loss event probability: $P_e = \frac{pq}{p + q}$
  - Average loss burst length: $\bar{n} = \frac{1}{q}$

![Diagram of Gilbert-Elliott Model](attachment:image.png)
Frame Complexity Example

[ (a) Low Frame Complexity ]

[ (c) High Frame Complexity ]

[ (b) Medium Frame Complexity ]
Contribution

- Proposed an effective way of assessing perceived video quality
- Reversed reduced-reference method reduces the amount of traffic for transmitting the source video to client
- Having analyzer at the server keeps client lightweight
- Keeping the causality among different layers makes a management easier
- **Subjective video quality metric** based on defective events better reflect human perception than simple image quality
- Event-based subjective quality metric enables measurement of user experience of a service