AI-Powered Per-Scene Live Encoding

Anita Chen | W3C Workshop on Web and Machine Learning | September 2020
Agenda

1. Per-Title Encoding: basics
2. Web-based AI Solution for Per-Title/Per-Scene Encoding
3. Summary and Outlook
Per-Title Encoding - Basics
PER-TITLE ENCODING | what? why?

<table>
<thead>
<tr>
<th>Resolution</th>
<th>Bitrate (kb/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>416x234</td>
<td>145</td>
</tr>
<tr>
<td>640x360</td>
<td>365</td>
</tr>
<tr>
<td>512x384</td>
<td>560</td>
</tr>
<tr>
<td>768x432</td>
<td>730</td>
</tr>
<tr>
<td>768x432</td>
<td>2000</td>
</tr>
<tr>
<td>1280x720</td>
<td>3000</td>
</tr>
<tr>
<td>1280x720</td>
<td>4500</td>
</tr>
<tr>
<td>1920x1080</td>
<td>6000</td>
</tr>
<tr>
<td>1920x1080</td>
<td>7800</td>
</tr>
</tbody>
</table>

Apple h264 Encoding Ladder

Low complexity / High redundancy
- Animation

High complexity / Less redundancy
- Action
- Nature documentaries
- Sport

Applied across all types of content
**Encode Complexity Analysis**  
- Perform test encodings with different settings and calculate corresponding VMAF values

**Convex Hull Estimation**  
- Select bitrate/resolution pairs that are close to the convex hull

**Production Encoding**  
- Perform production encoding using the optimal encoding ladder

Reference video: 1080p sports video

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Bitrate (kb/s)</th>
<th>VMAF</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>7800</td>
<td>99</td>
<td>42.1</td>
</tr>
<tr>
<td>Per-title</td>
<td>3370</td>
<td>94.5</td>
<td>37.6</td>
</tr>
<tr>
<td>Per-scene</td>
<td>2170</td>
<td>93</td>
<td>35.4</td>
</tr>
</tbody>
</table>
Web-based AI Solution for Per-Title/Per-Scene Encoding
AI-Powered PTE & PSE | workflow

**Source video**
Live & VoD

- Feed video source into distribution workflow

**Complexity Analysis**
- ...analyze source information (resolution, frame rate), calculate quality, review test encodings for reference

**Machine Learning**
- ...select appropriate regression algorithm based on complexity analysis to predict optimized encoding profile

**Encoding & Packaging**
- Encode video with optimized profile

**CDN**
- ...provide availability for playback streaming

**Video Player**
- Play optimized stream

- Return streaming metrics and QoS parameters
AI-Powered PTE & PSE | Usage of Web APIs

Source video
Live & VoD

- Video file upload (browser)
- Video upload via link (browser)

Complexity Analysis

- Server-side analysis
- Analysis data sent to model via endpoint
- Client-side analysis

Machine Learning

- Client-side predictions
- Model Loader & Web Neural Network API's
  - town hall meetings
  - live streaming scenarios

Encoding & Packaging

- Server-side production encoding
- Utilize WebCodecs API for town hall meetings

CDN

- Video Player
- Current
- Upcoming
AI-Powered PTE & PSE | web interface

Drag and Drop video file here or click the button to upload. Maximum allowed file size: 3.00 GB

Select Video

Task progress

<table>
<thead>
<tr>
<th>Video</th>
<th>Analysis</th>
<th>Characteristics</th>
<th>Calculation</th>
<th>Video Content</th>
<th>Scene Change</th>
<th>Detection</th>
<th>SI / TI</th>
<th>Scene</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marlbo-Channels.mp4</td>
<td>Started a few seconds ago</td>
<td>Started a few seconds ago</td>
<td>Started a few seconds ago</td>
<td>Started a few seconds ago</td>
<td>Started a few seconds ago</td>
<td>Started a few seconds ago</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Bitrate-Quality Comparison

- Apple h264 Encoding Ladder
- FAMiUM Predicted Encoding Ladder
- Calculated VMAF

Predicted Encoding Ladder

<table>
<thead>
<tr>
<th>#</th>
<th>Resolution</th>
<th>Bitrate</th>
<th>VMAF (predicted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>960 x 540</td>
<td>238</td>
<td>63</td>
</tr>
<tr>
<td>2</td>
<td>640 x 360</td>
<td>265</td>
<td>69</td>
</tr>
<tr>
<td>3</td>
<td>768 x 432</td>
<td>401</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>738 x 392</td>
<td>512</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>960 x 540</td>
<td>666</td>
<td>87</td>
</tr>
<tr>
<td>6</td>
<td>1280 x 720</td>
<td>891</td>
<td>93</td>
</tr>
<tr>
<td>7</td>
<td>1920 x 1080</td>
<td>2182</td>
<td>99</td>
</tr>
</tbody>
</table>
Stacked Model
- Framework: Keras
- flexible in terms of its concept in combining models
- each model requires several cycles in fine-tuning in order to perform well together

XGBoost
- Framework: SKlearn
- does not require as much normalization for training attributes
- certain encoding methods can weaken the model performance

Convolutional Neural Network
- Framework: Tensorflow
- flexible and video-friendly in terms of processing
- computationally expensive, training can be slow without a strong GPU

Feed Forward
- Framework: Keras
- robust enough to support missing input values
- requires a large amount of data for better performance
3

Summary & Outlook
Deep Encode

✓ No computationally heavy test encodes

✓ Metadata extraction and AI-based image processing for content analysis

✓ Deep Learning for appropriate encoding ladders
SUMMARY & OUTLOOK | what's next?

- Model Loader API:
  - Integrate into Solution
  - Live streaming & company-wide conference call use cases
  - Contribute to API to support Deep Encode on the Web

- Model format standardization contribution

- Utilize web codecs for live streaming scenario

- Prediction of other types of video quality

- Exploration of additional video feature detection algorithms