Smooth Streaming over Wireless Networks

By
Sreya Chakraborty
Under guidance of
Dr. K.R.Rao

April 28, 2011
Introduction

- Smooth streaming is a challenge in areas where bandwidth is low or limited.
- Modern video transmission and storage are based on RTP/IP for real time services.
- Most RTP/IP access networks are typically characterized by a wide range of connection qualities and receiving devices.
- The varying connection quality is due to adaptive resource sharing mechanisms of these networks.
Implications of video traffic smoothing on the numbers of statistically multiplexed H.264SVC, H.264/AVC and MPEG-4 part 2 streams [2]

Traditional digital video transmission and storage systems are based on H.222.0, H.320 [7]

International video coding standards H.262, H.263 and MPEG-4 [1],[2]

Simulcast provides similar functionalities as a scalable bit stream.
Effectiveness of two elementary techniques for mitigating high traffic variability

- Buffered multiplexing: Identify the multiplexer buffer sizes required to support maximum number of streams.

- Video traffic smoothing
From the wide spectrum of video traffic smoothing techniques two approaches are:

- **Optimal Smoothing**
  - minimizes the traffic variability.
  - Computational complexity of $O(M)$, $M$ is the number of video frames.

- **Basic Smoothing**
  - averages the sizes of a prescribed number of video frames.
  - Computationally simple $O(1)$
Fig. 1 Basic compression system [10]
Fig. 2 Typical coding system [10]
Fig. 3: Block diagram of H.264 [14]
Fig. 4: Luma prediction (intra-prediction) modes in H.264 [15]
Fig. 5: Macroblock portioning in H.264 for inter prediction (a) (L–R) 16x16, 8x16, 16x8, 8x8 blocks; (b) (L–R) 8x8, 4x8, 8x4, 4x4 blocks [15]
JM Software [12]:
This software is a product of Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG.
The latest version of JM Software is 17.2. It supports both planar and interleaved/packed raw image data (viz., yuv, rgb). The input file is a configuration file (text file) and some of the parameters passed in that file are:
- Input file
- Number of frames to be encoded
- Frame rate
- Output frame width and Height
- Profile, level selection
- GOP size
- Bit rate control
JSVM software [13]:
This software is a product of Joint Video Team (JVT) of ISO/IEC MPEG and ITU–T VCEG.
The input file is a configuration file (text file) and some of the parameters passed in that file are:
- Input file
- Number of frames to be encoded
- Frame rate
- Output frame width and Height
- Profile, level selection
- GOP size
- Bit rate control
JM(17.2) Performance Analysis

JM Performance in Baseline Profile
- Video Sequence – akiyo_qcif
- Number of frames encoded – 25
- GOP – IBPBPB
- Quantization parameter – 25, 30, 35, 40
- Number of reference frames – 3
Video Sequences Used:
- Baseline
- File Size: 3713 KB
- QCIF format – 176 x 144
- YUV – 4:2:0
<table>
<thead>
<tr>
<th>Quantization Parameter</th>
<th>Peak to Peak Signal to Noise ratio (dB)</th>
<th>Total Encoding time(s)</th>
<th>Bitrate (Kbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>51.967</td>
<td>160.654</td>
<td>4333.35</td>
</tr>
<tr>
<td>20</td>
<td>44.576</td>
<td>149.769</td>
<td>1547.8</td>
</tr>
<tr>
<td>25</td>
<td>41.26</td>
<td>141.364</td>
<td>884.57</td>
</tr>
<tr>
<td>30</td>
<td>37.606</td>
<td>131.664</td>
<td>455.68</td>
</tr>
<tr>
<td>35</td>
<td>34.968</td>
<td>120.128</td>
<td>241.37</td>
</tr>
<tr>
<td>40</td>
<td>32.957</td>
<td>107.499</td>
<td>132.43</td>
</tr>
</tbody>
</table>

Table 1: JM performance in Baseline profile
Fig. 6: PSNR vs QP for Baseline profile of akiyo_qcif
Fig. 7: Encoding time vs QP for baseline profile of akiyo_qcif
Scalable Video Coding
Application scenario of SVC

[Diagram showing the flow of a scene through an SVC encoder, with branches for different bitrates (128 kbit/s, 256 kbit/s, 512 kbit/s, 1024 kbit/s) leading to H.264/AVC decoders, and further splitting to SVC decoders with resolutions QCI, CIF, TV at 7.5 Hz, 15 Hz, and 60 Hz, respectively, connected to a video server and network.]
What is H.264 / MPEG-4 SVC

- Scalable Video Coding (SVC) will be an extension of the H.264 / MPEG-4 Advanced Video Coding (AVC) standard
  - Starting point of SVC extension was designed and proposed by Fraunhofer Institute HHI
  - Standardization of SVC extension is still under way
Types of scalability

Temporal Scalability

30 fps
15 fps
7.5 fps

Spatial Scalability

QCIF
CIF
TV

Quality Scalability
To serve different needs of different users with different displays connected through different network links by using a single bit stream, i.e., a single coded version of the video content:

- Spatial scalability: Choose appropriate resolution
- Temporal scalability: Choose convenient frame rate
- Quality scalability: Choose suitable data rate
  by removing parts of the bit stream
Fig. 8: B frame prediction structures [8]
• Hierarchical B frame is used.

• For the H.264 SVC encodings GoP structure IBBBBBBBBBBBBBBBBBB (16 frames, 15 B frames per I frame) denoted by G 16–B15

• Statistical video traffic analysis demonstrates that these encoding parameters settings and GoP structures result overall in a very good rate distortion efficiencies for the respective encoders.
The SVC extension is built on H.264 / MPEG-4 AVC and re-uses most of its innovative components. As a distinctive feature, SVC generates an H.264 / MPEG-4 AVC compliant, i.e., backwards-compatible base layer and one or several enhancement layer(s). The base layer bit stream corresponds to a minimum quality, frame rate, and resolution (e.g., QCIF video), and the enhancement layer bit streams represent the same video at gradually increased quality and/or increased resolution (e.g., CIF) and/or increased frame rate.
Fig. 9: Block diagram of SVC [1]
Fig. 10: SNR Scalable

Base layer

Upper layer
Fig. 11: Decoding process for SNR scalability
What are the benefits in terms of applications

- Ease of adaptation to different terminal capabilities
- Resource conserving transmission, storage, and display of video, e.g., in surveillance applications
- Higher transmission robustness, if combined with unequal error protection
- Ease of Multicast Streaming through heterogeneous networks
Incorporates multiple streams in a single stream.

Customized

Can send a single video stream to multiple heterogenous clients

Bandwidth and storage space is saved. The base video stream layer of lower quality can be stored separately, instead of storing all the layers. This might be useful for video surveillance

H.264 SVC can give a decent (manageable) picture quality even at 20–40% packet loss in the network while the maximum tolerable packet loss for H.264 AVC might be around 1–5%.
• H.264 SVC [1] video encoding is expected to be widely adopted for wired and wireless networks video transport due to their compression efficiency.

• SVC enables the transmission and decoding of partial bit streams to provide video services with lower temporal or spatial resolutions.

JSVM Performance in Baseline Profile
- Video Sequence – Die Hard
- Number of frames encoded – 30
- GOP – G16B15
- Quantization parameter – 25, 30, 35, 40
General Information

Sequence: Die Hard
Resolution: 352x288
FPS: 30.0
Frames: 53984
Encoder: JSVM
Encoding type: High (Level 2.2), Scalable High (Level 3.1)
GoP pattern: G16B15
Other: 1,2,2,3,4,4

Fig. 12: Video sequence Die Hard [11]
Trace preview for the video sequence Die Hard [11]

<table>
<thead>
<tr>
<th>#</th>
<th>Frame Time [s]</th>
<th>Type</th>
<th>Size [Byte]</th>
<th>PSNR-Y [dB]</th>
<th>PSNR-U [dB]</th>
<th>PSNR-V [dB]</th>
<th>Priority (63..0 if frame present, -1 otherwise)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>I</td>
<td>53</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>1</td>
<td>0.03333</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>2</td>
<td>0.06667</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>0.1</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>4</td>
<td>0.13333</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>0.16667</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>6</td>
<td>0.2</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>7</td>
<td>0.23333</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>8</td>
<td>0.26667</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>9</td>
<td>0.3</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>10</td>
<td>0.33333</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>11</td>
<td>0.36667</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>12</td>
<td>0.4</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>13</td>
<td>0.43333</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>14</td>
<td>0.46667</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>15</td>
<td>0.5</td>
<td>B</td>
<td>13</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
<tr>
<td>16</td>
<td>0.53333</td>
<td>I</td>
<td>53</td>
<td>188.1308</td>
<td>188.1308</td>
<td>188.1308</td>
<td>63</td>
</tr>
</tbody>
</table>
Fig. 13. Peak/Mean of size vs Average quality (PSNR-Y) for Die Hard [11]
Fig. 14: Average quality (PSNR–Y) vs Average bit rate for Die Hard [11]
**General Information**

<table>
<thead>
<tr>
<th>Sequence: Citizen Kane</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution: 352x288</td>
</tr>
<tr>
<td>FPS: 30.0</td>
</tr>
<tr>
<td>Frames: 53984</td>
</tr>
<tr>
<td>Encoder: JSVM1</td>
</tr>
<tr>
<td>Encoding type: High (Level 2.2), Scalable High (Level 3.1)</td>
</tr>
<tr>
<td>GoP pattern: G16B15</td>
</tr>
<tr>
<td>Other: 1,2,2,3,4,4</td>
</tr>
</tbody>
</table>

**Fig.15: Video sequence Citizen Kane [11]**
Fig. 16: Peak/Mean of size vs Average quality (PSNR-Y) for Citizen Kane [11]
Fig. 17: Average quality (PSNR–Y) vs Average bit rate for Citizen Kane [11]
Conclusions:

- In comparison to the scalable profiles of prior video coding standards, the H.264/AVC extension for scalable video coding (SVC) provides various tools for reducing the loss in coding efficiency relative to single-layer coding. The most important differences are:

  1. The possibility to employ hierarchical prediction structures for providing temporal scalability with several layers while improving the coding efficiency and increasing the effectiveness of quality and spatial scalable coding.

  2. New methods for inter-layer prediction of motion and residual, improving the coding efficiency of spatial scalable and quality scalable coding.

  3. The concept of key pictures for efficiently controlling the drift for packet-based quality scalable coding with hierarchical prediction structures.
(4) Single motion compensation loop decoding for spatial and quality scalable coding providing a decoder complexity close to that of single-layer coding.

(5) The support of a modified decoding process that allows a lossless and low-complexity rewriting of a quality scalable bit stream into a bit stream that conforms to a non-scalable H.264/AVC profile.

- These new features provide SVC with a competitive rate-distortion performance while only requiring a single motion compensation loop at the decoder side.

(1) Temporal scalability: can be typically achieved without losses in rate-distortion performance.
(2) Spatial scalability: It should be noted that the results typically become worse as spatial resolution of both layers decreases and results improve as spatial resolution increases.

(3) SNR scalability: when applying an optimized encoder control, the bit rate increase relative to non-scalable H.264/AVC coding at the same fidelity can be as low as 10% for all supported rate points when spanning a bit rate range with a factor of 2–3 between the lowest and highest supported rate point.
References


