MM-1 Architecture of a Quality Based Intelligent Proxy (QBIX) for MPEG-4 Videos

Paper Review

By Ken Schmidt

Objectives of Project

- Create an intelligent proxy server to:
  - Allow clients with reduced capabilities to view MPEG-4 & MPEG-7 videos who could not currently view them
  - Reduce latency
  - Develop in an open source atmosphere

Introduction

- Web based video objects are large and will become larger (longer, higher quality)
- Video objects will become more frequently used (distance education, digital libraries, video on demand, etc)
- Internet/internet bandwidths may become insufficient & download latencies may become excessive
- Conventional caching of web objects is not appropriate for large videos (cache replacement)
- Use a Quality Based Intelligent Proxy (QBIX) to cache, compress (MPEG4) adaptively to user capabilities (with limited bandwidth, display or CPU capabilities)

Background

- The number of video streams is increasing, video (and audio) is much larger than conventional Internet traffic (even when compressed)
- Video streaming could easily outstrip the capacity of the Internet
- The quantity of video that can become available is huge
  - Video archives on line
  - There is a distribution in demand, there is not just one movie that is popular
  - caching the few movies that will fit into cache will not reduce bandwidth, each one will likely have to be replaced before it is reused from cache
  - Adding to cache size is not practical as the use of video will continue to grow

Background (cont)

- New methods need to be implemented for caching video
- New methods need to be implemented so that low bandwidth, slow CPU clients can view video
- New methods need to be implemented so that cache sizes may be reduced for streaming to small, low resolution screens (PDAs) or black and white screens
- Other video caching techniques have been studied:
  - Full and Partial video caching
  - Caching in the Time and Quality domain

Principle Contribution

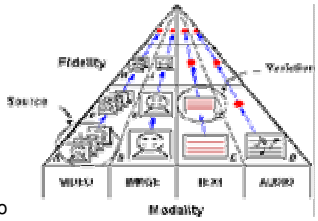
- This paper contributes an open source method of video caching in the quality domain
  - Using multimedia standards MPEG-4 and MPEG-7
  - And communication standards RTP, RTSP
- Only one other quality domain video caching proxy exists but relies on proprietary systems and protocols
- Partial video caching in the time domain presents problems for whole videos
  - Synchronization overhead (part at proxy, part at server
  - Interactivity delays
  - Jump from prefix to suffix causing an intermission

Techniques – What Existed

- MPEG-4 consists of a set of scenes, each containing a set of media objects which have
  - Position
  - Size
  - Shape
- Each object has at least one Elementary Stream (ES)
- MPEG-7 (June 2002) added to MPEG-4:
  - Description Definition Language (DDL) (allows DS)
  - Description Schema (DS) (used to code Descriptors)
  - Descriptors (describes audio & video features – color histogram, shapes, semantic text info-meta data)
  - They use meta data

Techniques – What Existed

- A client's capabilities may allow only one still image to be displayed rather than a movie
- Or a client may only be capable of displaying a text description of the video
- Or an image with the audio portion of the video
- Or an audio rendering of the text description



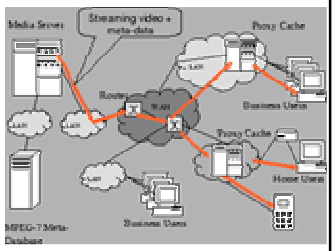
Techniques – What Existed

- CC/PP – Composite Capability/Preference Profiles – is an extension of HTTP 1.1
  - Meta data
    - Hardware block – display size, bandwidth, CPU, memory
    - Software block – OS, software capabilities (HTML version, sound support, etc), can an image be displayed, etc

```
<rdf:Description about="HardwarePlatform">
   <prf:Defaults
   Vendor="Nokia"
   Model="666"
   Type="PDA"
   ScreenSize="320x200x16"
   CPU="PPC"
   Keyboard="Yes"
   Memory="16MB"
   Speaker="Yes" />
</rdf:Description>
```

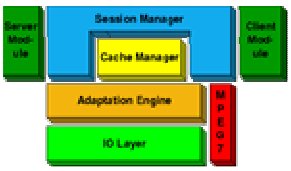
Techniques – Architecture

- Client sends request to proxy
  - Includes capabilities
- Proxy either
  - If a copy is available
    - Scales to user capabilities
      - Temporal domain
      - Color domain
      - Spatial domain
      - Bitrate domain
  - If a copy is not available
    - Requests a copy or
    - Requests a pre scaled copy



Techniques – Architecture of Proxy

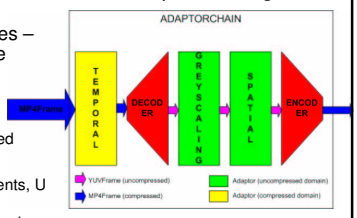
- Session Mgr:
  - Server Mod: imitates media server for client
  - Client Mod: imitates client for media server
- Cache Mgr: caches videos and cache replacement strategies
- MPEG7 Mod: parse and generate MPEG7 descriptions
- Adaptation Eng: scales video to user capabilities



IO Layer: Hides network and file access behind one abstract class

Techniques – Architecture of Adaptation Engine

- Temporal: drops frames – to avoid artifacts in the video:
  - 1st drop all B frames
  - 2nd drop P frames
  - I frames are not dropped
- Color:
  - Chrominance components, U & V, set to gray scale
  - Bitrate scaled for reduced color information
- Spatial: down sample the picture to the spatial size required by the client – uses nearest neighborhood algorithm – the simplest, calculates quickly.



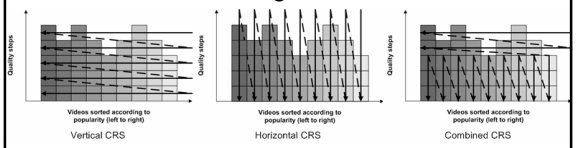
- i-frame = "intra-frame": basic type stores all of the data required to display the frame
- p-frame = "predictive frame": built from a previous i or p-frame; differences between frames
- b-frame = "bidirectional frame": built from two p-frames, can predict future frame from previous frame, combine a past and future frame

Techniques – Architecture of MPEG-7 Module

- Creates meta data for MPEG-4 video
  - Size
  - Type (video or audio)
  - Bit rate for each elementary stream (object in the scene)
- Variation descriptor contains
  - Name of the adaptation step
  - Expected quality
  - Loss
  - Priority of adaptation step
- Example:
  - Fidelity 75% (loss 25%
  - Size saved = 6.25 Mbytes (1/2)
  - Bit rage halved

```
<Description xsi:type="VariationDescriptionType">
 <VariationSet>
  <Source xsi:type="VideoType"> [..]
   <ComponentMediaProfile id="ES1"> [..]
   <ComponentMediaProfile id="ES2"> [..]
   <MediaFormat>
    <Content href="MPEG7ContentCS">
     <Name>audiovisual</Name>
    </Content>
    <FileFormat
     href="urn:mpeg:mpeg7:cs:FileFormatCS:2001:5">
     <Name xml:lang="en">mp4</Name>
    </FileFormat>
    <FileSize>13107200</FileSize>
    <BitRate>131072</BitRate>
   </MediaFormat>
  [..]
  </Source>
  <Variation id="VARIATION1"
          fidelity="0.75"
          priority="1"> [..]
  <Variation id="VARIATION2"
          fidelity="0.45"
          priority="2"> [..]
 </VariationSet>
</Description>
```

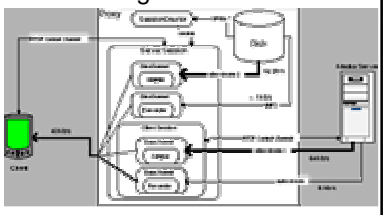
Techniques – Architecture of Cache Management Module



- Vertical CRS – successively chooses quality variations of least popular video in the list for replacement, degenerates to LRU (Least Recently Used)
- Horizontal CRS – chooses adaptation candidates according to their quality, tends to adapt every object down to its lowest quality layer, even for newly inserted videos
- Combined CRS – overcomes disadvantages of Vert and Horz
- Currently, if more than one variation, first set is chosen
- Future work
  - Compare to sets and decide which one to use
  - Consider adaptation costs

Techniques – Architecture of Session Management

- Server session connects client to proxy
- Client session connects proxy to server



Example: 4 ES – 2 video, 1 audio, 1 BIFS. Client can accept 40 kbps (tells this to Proxy). Original video is 136 kbps, media server does not have a lower bandwidth version. Video has been cached before but only BIFS and one video stream remains in cache. Session Creator reads MPEG7, recognizes a mismatch, returns sequence of adaptation steps. Media server streams the missing video stream and audio. Each video stream (one from cache, one from server) is adapted from 64 kbps to 16 kpbs (bitrate scaling), audio is not scaled. Total bandwidth is now 40 kbps.

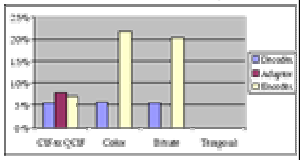
Testing Performed on Adaptors

Computational performance of implemented adaptors was tested, only slight deviation, 1-2%, occurred for different video content. Example:

- Dimension: 352 x 288 CIF
- Frame rate: 25 fps
- Duration: 23 min, 44 sec
- Bitrate: 1 Mbit/sec
- 1.6 GHz Pentium IV
- 512 MB RAM
- Windows 2000
- Decoding not needed for Temporal
- Bitrate scaling does not use an adaptor, done in encoder
- CIF to QCIF downsamples to 176 x 144 (spatial scaling)

	CIF to QCIF (Spatial)	Color	Bitrate	Temporal
Decoder	79779	79983	77100	0
Adaptor	113028	281	0	31
Encoder	101405	308183	289759	0
Total	294212	388447	366859	31

Time (msec) for different adaptors



Adaptation time, % of total video playback length

Advantages & Disadvantages of this approach

- Advantages
  - Allows clients with reduced capabilities to view videos
  - May reduce startup delay
    - For full capacity clients
    - For more clients if reduced videos are cached
  - Is open source
    - many would be willing to make improvements
    - License cost free to implement
- Disadvantages
  - Adaptation is very costly
    - Currently limited to 3 simultaneous clients using adaptation
    - Scalability is severely limited
      - Develop adaptors that work in the compressed domain
      - Content providers offer video streams with native scalability options
      - Use distributed proxy architecture or content distribution network

Future Work

- Develop adaptors that work in the compressed domain
- Improve LRU cache replacement with
  - Intelligent proxy which compares two sets to decide which one to use
  - Develop more advanced cache replacement strategies to consider adaptation costs

Critical Evaluation

- While the paper suggested that adaptation could be scaled from 3 simultaneous clients, it indicated that this could be overcome using 3 techniques. Even if scalability were increased by a factor of 10, the user/proxy ration will still not be cost effective
- There was no indication of who would be willing to pay for proxies (who would gain something from the expenditure on proxies)
- Overall, this paper does a good job describing advantages and limitations of the project

Quiz

1. Which adaptation technique does not require decoding/encoding?
   a. Spatial b. Color c. Temporal
2. What does the Proxy I/O layer do?
3. The Session Management Server session connects ____ to ____
4. T or F: The spatial adaptation engine drops P frames first
5. T or F (if F, enter the correction) MPEG-7 provides more highly compressed video than MPEG-4 video_____

Quiz Key

1. c) Temporal adaptation does not require an encoder/decoder
2. The Proxy I/O layer hides network and file access behind one abstract class
3. The Session Management Server session connects client to proxy.
4. False (for more than one reason), the spatial adaptation engine downsamples the pixels using the nearest neighborhood. The Temporal engine drops B frames first.
5. False: MPEG-7 provides meta data about an MPEG-4 video