UltraGrid: Low-Latency High-Quality Video Transmissions on Commodity Hardware

Petr Holub
CESNET z.s.p.o., Prague/Brno, Czech Republic

<Petr.Holub@cesnet.cz>

Network Performing Arts Production Workshop
2013–03–12
UltraGrid Platform

- Technology
  - an affordable platform for high-quality interactive image transmissions
  - use of commodity hardware
    - Linux PC and Mac platforms
    - commodity video capture cards
    - commodity GPU cards
    - 10GE is a plus but not necessary
  - as low latency as possible on commodity hardware
  - open-source software, BSD license
  - a platform for implementing research results (not just ours! :)
    - compression & image processing, FEC, scheduling, congestion control...
Applications of UltraGrid

- Generic scientific visualization
- Medicine
  - X-ray imagery, cardiology, pathology
Applications of UltraGrid

- Education
  - remote education
Applications of UltraGrid

- Cinematography

Detached BaseLight consoles at CinePost (Barrandov, CZ)

Mac Pro

Kona 3

dual-link HD-SDI

BaseLight Four

SONY SXRD 4K

dual-link HD-SDI

10GbE
Applications of UltraGrid

- Arts
  - distributed performances: music, theater
UltraGrid Platform

- History of Development
  - 2005–now: CESNET (→ 1080i)
  - 2006–2008: forks by KISTI (AJA KONA) and i2cat (SAGE)
  - 2012–now: i2cat (H.264)

- Some milestones
  - 2002: 720p
  - 2005: 1080i, multipoint
  - 2007: CPU compressions, self-organization, optical multicast
  - 2008: 2K/4K
  - 2011: GPU compressions
  - 2012: 8K
UltraGrid Platform

- Supported formats
  - HD, 2K
  - 4K – tiled or native
  - 8K – new
  - multichannel video (e.g., 3D HD, 4K)
- Uncompressed vs. compressed
  - low-latency compression
  - GLSL-accelerated DXT1, DXT5-YCoCg
  - CUDA-accelerated JPEG, DXT5-YCoCg
  - CPU-based DXT1, ffmpeg (e.g., H.264)
- Supported audio formats
  - uncompressed, multi-channel
## UltraGrid Platform

### I/O
- Capture/playback cards: HD-SDI, SDI, HDMI, analog HD and SD
  - Manufacturers’ SDKs, Video4Linux2, QuickTime
- Screen capture input
- Computer screen output (OpenGL, SDL)
- SAGE output
- Specialized display filters
- Stereoscopic HDMI 1.4a

### Full-duplex operation

### Simple GUI
- QT-based, native MacOS
- Permanent storage of configuration
- Simple startup + advanced configuration dialog

Line-interlaced stereoscopic video
UltraGrid Platform

GUI on MacOS X
UltraGrid Platform

GUI on Linux
Audio

- balanced, unbalanced, HD-SDI, HDMI
- various system interfaces including JACK
- PortAudio, ALSA, CoreAudio, JACK
- embedded HD-SDI/HDMI
- simple mono software echo canceler based on Speex
- channel mixer/duplicator
• Available compression schemes
  - DXT1: CPU-based (FastDXT library from EVL)
  - DXT1, DXT5: OpenGL Shader Language (GLSL) based
  - JPEG: NVidia CUDA based
  - DXT5: NVidia CUDA based (for 8K)

SAGE display with various compressions
GPU-Accelerated Compression

- Fine-grained parallelization of JPEG
  - per-row/column DCT/IDCT
  - per pixel RLE
  - per pixel Huffman
  - parallel stream compacting
  - parallel decompression using restart intervals
  - use of auxiliary indexes for more efficient parsing

- Available also as BSD-licensed open-source library:
  http://gpujpeg.sf.net/
GPU-Accelerated Compression

- Fine-grained parallelization of JPEG

```
tmp = __clz(map & mask); pzc = 2*(tmp - (32 - tid)); if ((0x80000000 >> tmp) > (map_o & mask)) {pzc++;}
pzc = 0 0 0 0 1 0 1 0 1 3 0 0 1 3 0 0 0 0 0 0 0 0 2 4 6 8 10 12 14 16 18 20 22
Decompose to zeros before even and odd elements.
```

Figure 3: Efficient formulation of sample-parallel RLE for GPUs. Note the condition following clz: due to working with pairs merged by OR operation, we need to distinguish whether the first non-zero pair is actually composed of both even and odd non-zero elements or not. If the odd element is zero, the count of preceding zeros, pzc, needs to be incremented by one.
GPU-Accelerated Compression

- Performance numbers (including transfer to/from GPU)
  - DXT1 GLSL: 798 Mpix/s (NVidia 580GTX), 593 Mpix/s (ATI 6990)
  - DXT5 GLSL: 349 Mpix/s (NVidia 580GTX), 305 Mpix/s (ATI 6990)
  - JPEG CUDA: up to 1.580 Mpix/s = 4.740 MB/s (NVidia 580GTX, 4:4:4, Q=60)
  - DXT5 CUDA: ≥1.580 Mpix/s (NVidia 580GTX)
GPU-Accelerated Compression

- JPEG performance

(a) Encoder performance (GPU only)
(b) Decoder performance (GPU only)
(c) Encoder performance (both CPU and GPU)
(d) Decoder performance (both CPU and GPU)
GPU-Accelerated Compression

- Performance of JPEG stages for 2160p video

![Graphs showing performance of JPEG stages](image)

(a) for JPEG encoder

(b) for JPEG decoder

Table 1: DXT compression performance. 4320p results are not available for ATI cards because of texture size limitation (4096×4096) of the drivers.
Forward Error Correction

- LDGM
  - CPU and GPU implementations
  - CPU (SSE optimized) is used because of CPU↔GPU transmissions overhead
  - packet loss up to 10% can be mitigated with reasonable overhead
  - can make JPEG survive up to 25% packet loss
- Simple method: shifted multiplication
Latency

- Latency limits
  - <150 ms for interactivity: ITU-T rec G.114

- End-to-end latency
  - in a local network
  - measured using video (1/60 s quantization)
  - depends substantially on hardware cards used (2.0–5.0 frames)
  - Bluefish444 should get us much lower: line-by-line API for HD-SDI
  - application-level traffic shaping to control bursts

- Uncompressed for DeckLink HD → DeltaCast 3G
  - 2.5 frames (83 ms)

- Impact of compressions
  - 2.5 frames (+<16.7 ms) for CUDA JPEG
  - 3.5 frames (+33.3 ms) for GLSL DXT1/5
User-Empowered Multi-Point Distribution

- UltraGrid supports multicast, but...
  - how available/dependable it is?
- UDP packet reflectors
  - controlled by the user
  - lower efficiency
  - possible per-user processing: transcoding, security,…
- Self-organization of the network
  - scheduling streams with bitrates comparable to capacity of links
  - CoUniverse framework (http://couniverse.sitola.cz)
  - constraints, MIP, local search
Users Worldwide

- source, binaries (http://ultragrid.sitola.cz/)
- embedded in SAGE (http://www.sagecommons.org/)
- Czech Republic (universities and university hospitals), USA (UCSD, UMich, UIC, Internet2, NLM/NIH, NorthwesternU, ...), Spain (i2cat, UPM), Portugal (FCCN), Netherlands (SARA), Poland (PSNC), Korea (KISTI), Russia, ...
Recent Updates
Since November 2012

- ffmpeg support – low latency H.264
  - 150% CPU core for HD, well usable at >18 Mb/s
  - 4K being examined
  - due to licensing issues, we don’t interface directly to X264 and leave it up to the user (GPL is viral and would propagate upstream)

- Windows port (almost done)
  - OpenGL, SDL displays
  - native BlackMagic SDK
  - DirectShow capture
Recent Updates
Since November 2012

- Support for DELTACAST DVI-I/DVI-D grabbers
  - ideal for content capture, computer screen resolutions
  - supports multiple cards (e.g., 6x DVI-I in in a single PC)
- File-based I/O
  - input/output of raw data
  - can be piped into mencoder (but not very convenient)
  - planned integration with further processing (e.g., GStreamer) for lecture/event/experiment recording, etc.
- Transcoding reflectors
  - change of formats “along the way”, as a part of multi-point data distribution
  - implemented using UltraGrid as backend
  - intended for automated setup with CoUniverse (later in 2013)
Recent Updates
Since November 2012

- Integration of 2-camera GColl
  - group-to-group communication with partial gaze awareness
Future Plans

- Short-term:
  - Advanced multi-point with scheduling (release with CoUniverse)
  - Software processor for multi-channel video

- Long-term:
  - Acceleration of low-latency H.264/H.265
  - New compression formats for specific purposes (e.g., SAGE)
World Firsts...

- 2005 – Multi-Point Uncompressed HD
  - n-way using packet reflectors

![Map showing connections between Brno, Chicago, San Diego, and Baton Rouge]
World Firsts…

- 2007 – Self-Organizing Multi-Point Uncompressed/Compressed HD
  - with CoUniverse
  - self-organizing multi-point distribution setup with uncompressed/DXT1 compression switching based on available bandwidth
World Firsts...

- 2011 – GPU-JPEG Transatlantic 4K

100 – 500 Mbit/s stream for HD/2K

- CineGrid Workshop, December 2011
- real-time movie post-production review/approval process
- playback on a machine worth $1,000 ($500 PC + $500 NVIDIA 580 GTX)
World Firsts...

- 2012 – GPU-JPEG Transatlantic Multi-Point 8K
  
  - from pre-rendered sources
  - JPEG → DXT5-YCoCg on a single machine
  - useful also as $16 \times$ HD (multi-camera setups)
Selected Papers


Thank you for your attention!

<petr.holub@cesnet.cz>
<ultragrid-dev@cesnet.cz>

This work is supported by LM2010005 project.