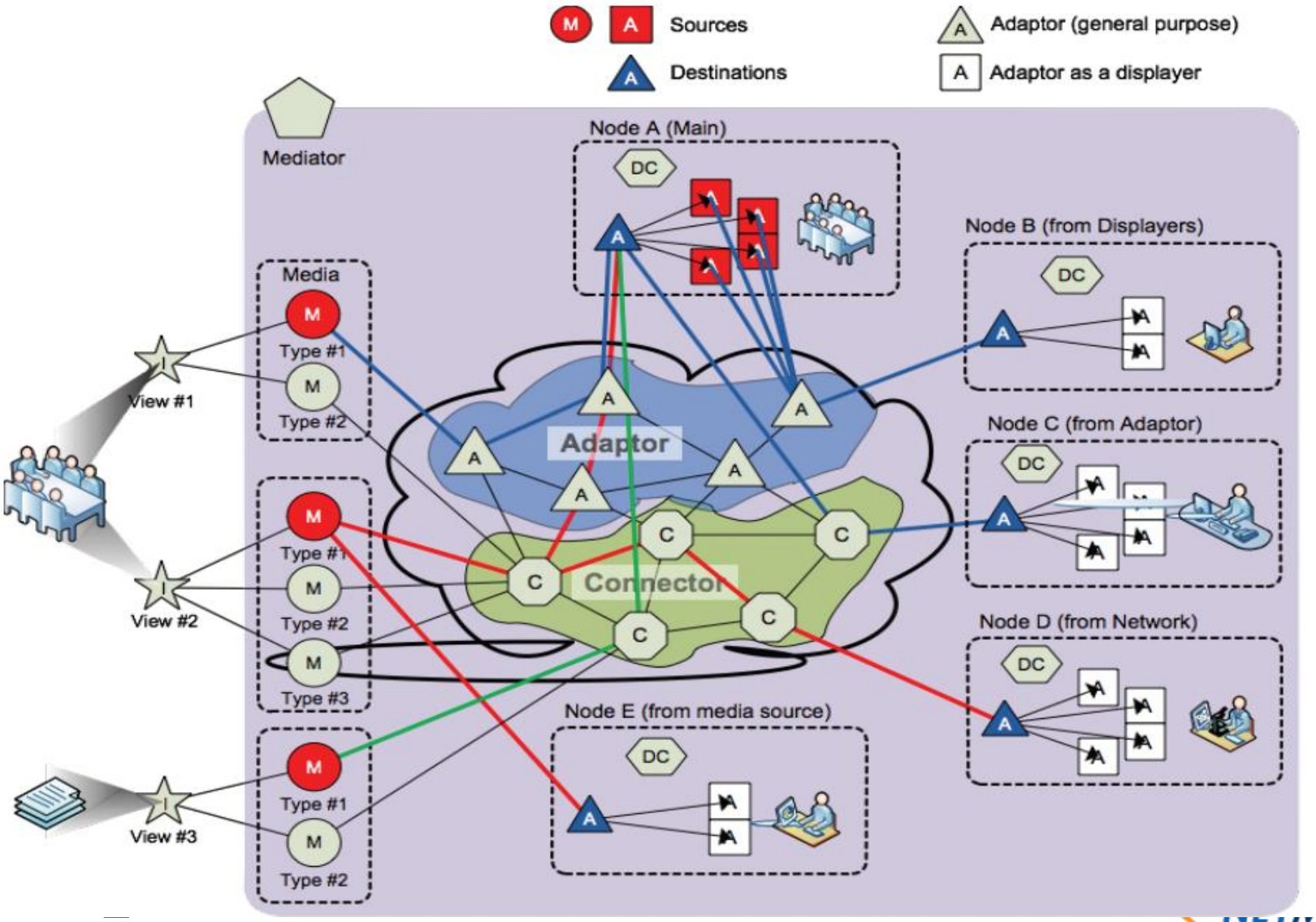
CUDA-supported real-time DXT compression of HD video: design and implementation

Feb. 23rd, 2011

HDTV WG Session of 31st APAN Hong Kong Meeting
@ Hong Kong, China

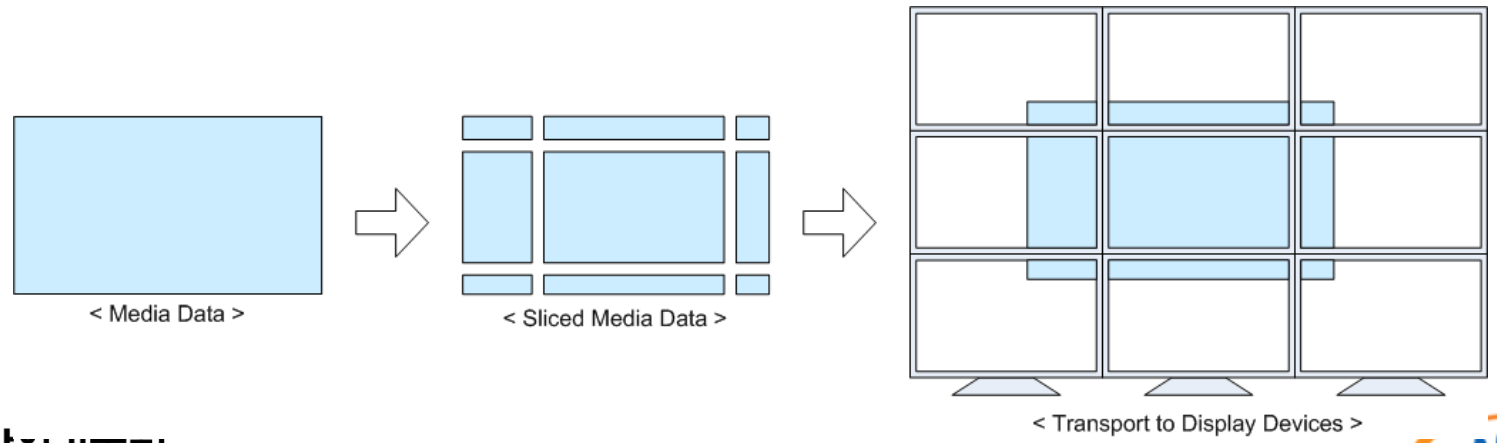
Dr. JongWon Kim
{jongwon}@nm.gist.ac.kr
Networked Media Lab., School of Information and Communications,
Gwangju Institute of Science and Technology (GIST), KOREA

Abstraction for Multi-party Visual Sharing



NeTD: Networked Tiled Display

- A networked display system using multiple tiled display devices to form a large logical display wall
- Provides ultra-high resolutions and large physical display sizes
- SAGE: Scalable Adaptive Graphics Environment
  - An "Operation System" for tiled-display environments
  - Manages the parallel graphics streams between the rendering nodes and tiled display nodes

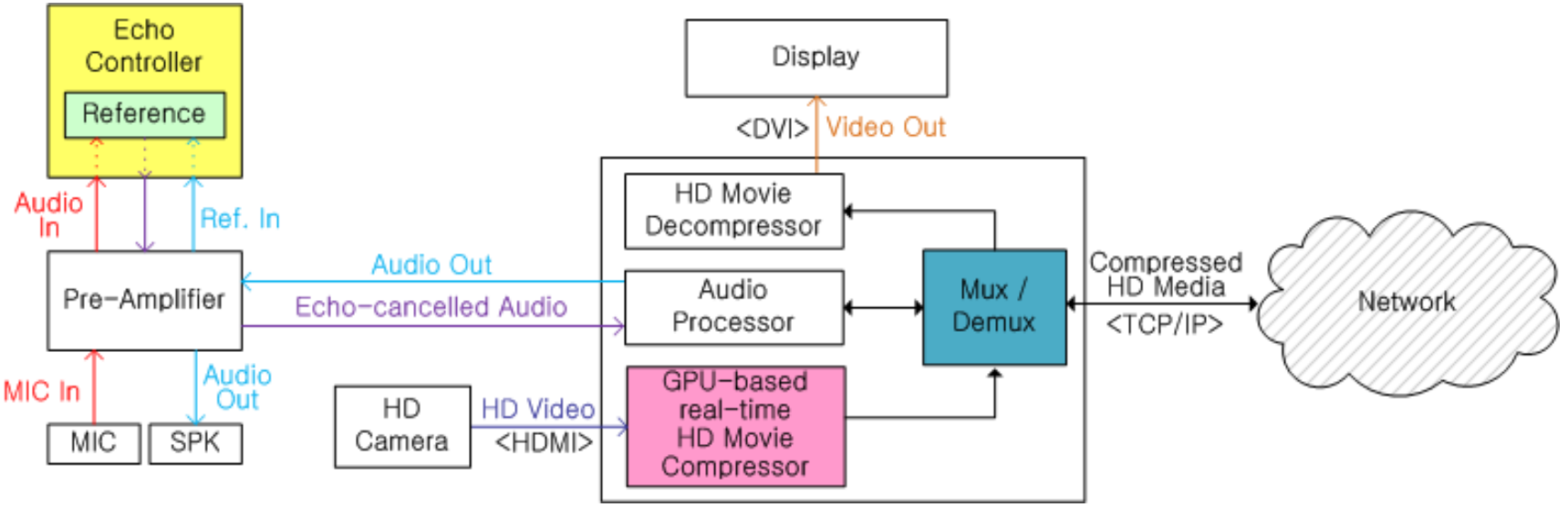


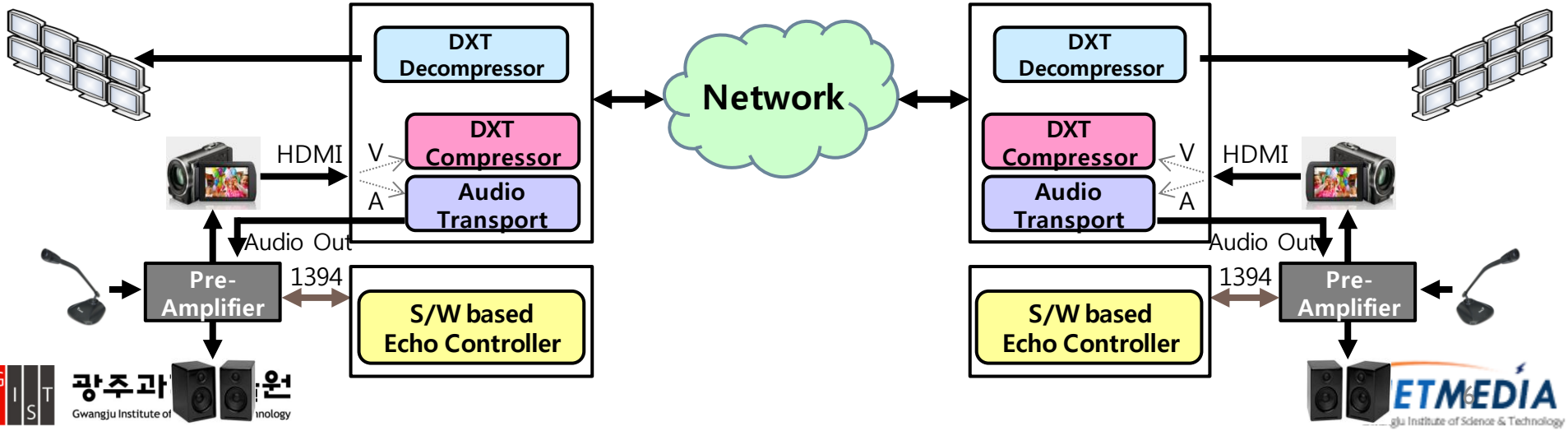
SAGE Visualcasting for Multi-party Collected Collaboration

HD Media Compression

- Many codecs for HD Media Compression
  - MPEG2, MPEG4, H.264, etc.,
  - Hard to realize real-time support
- DXT
  - S/W based light-weight lossy compression algorithm
  - Most of VGAs support H/W accelerated DXT-decompression
  - Each pixel block is independent of other blocks

HDMI-based HD Media Transport System



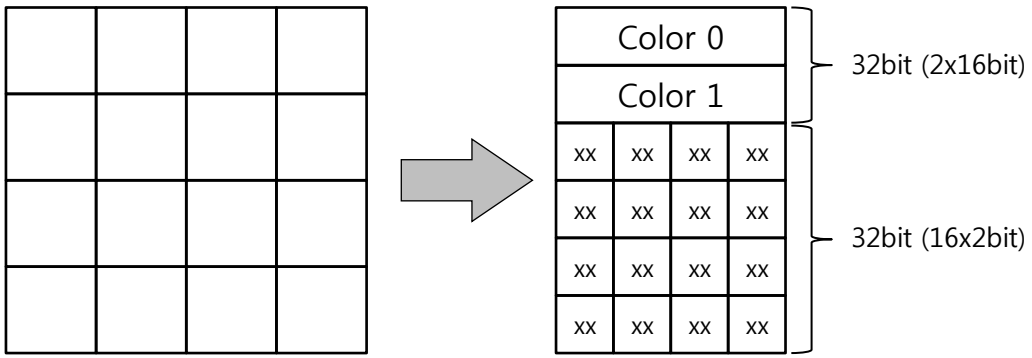


DXT: DirectX Texture Compression

- 4x4 block of pixels (512-bit or 384-bit) to a 64-bit or 128-bit quantity

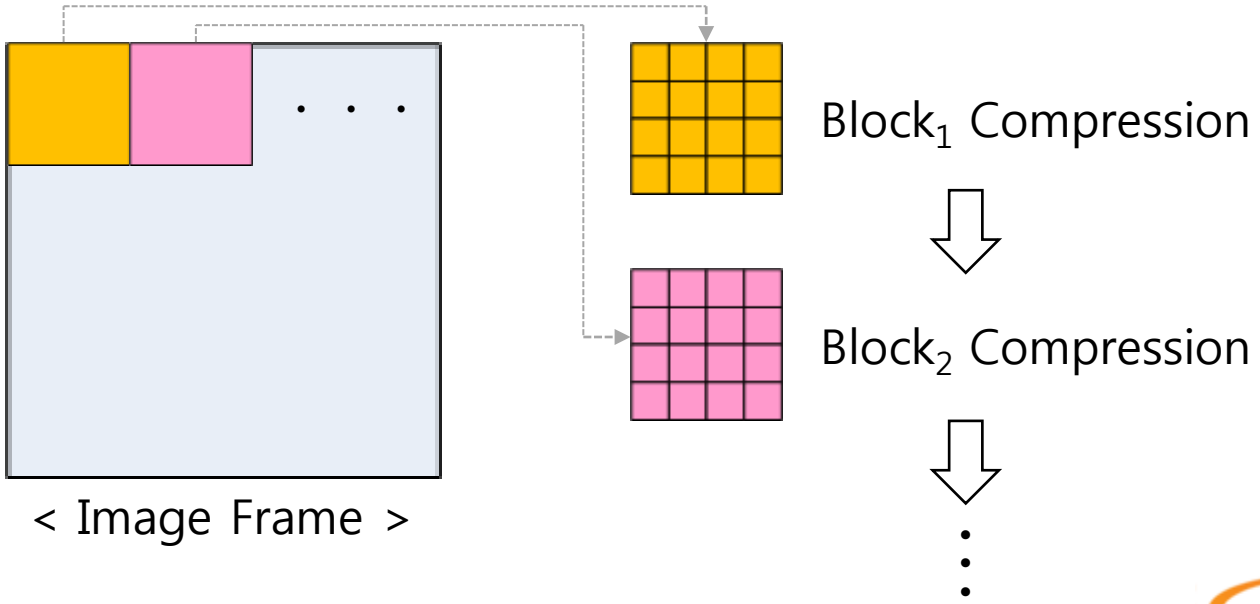
FOURCC	Description	Comp. ratio	Texture
DXT1 (BC1)	1-bit Alpha / Opaque	8:1 or 6:1	Simple non-alpha
DXT3 (BC2)	Explicit alpha	4:1	Sharp alpha
DXT5 (BC3)	Interpolated alpha	4:1	Gradient alpha

- Select DXT1 for HD Media Compression
  - HD Media does not include alpha channel

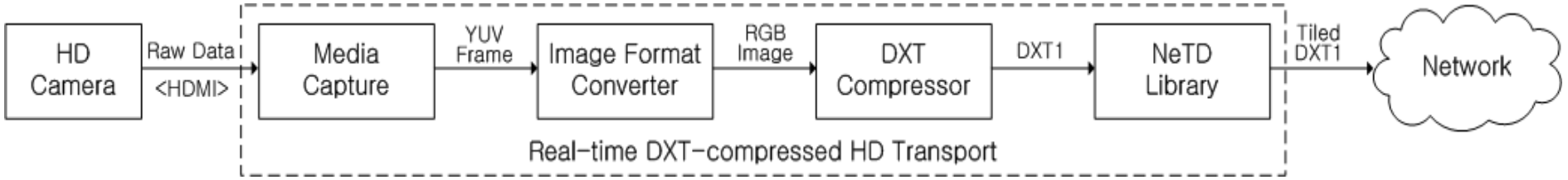


FastDXT: Realtime DXT Compression

- Focused on compression speed rather than quality
- Using Multi-Threads (2 or 4) for DXT Compression
- Optimized with SSE2 instruction-set
- Sequential compression processing each 4x4 pixel block



Performance of FastDXT for SAGE



- Using Decklinkcapture (HDMI Transport Application) by VisLab@UQ
- Performance

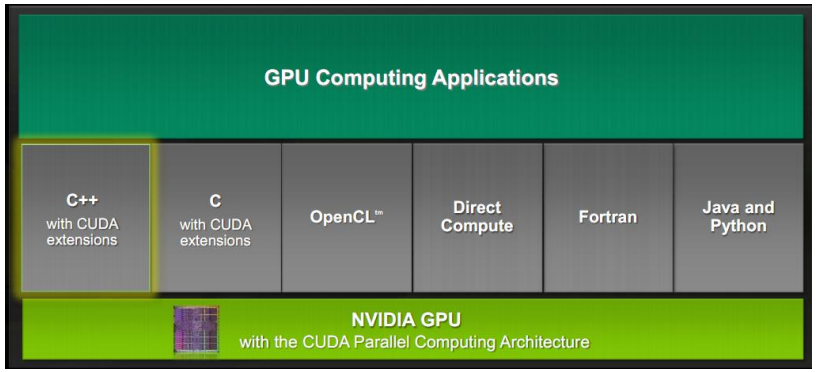
Machine	O/S	Compression	fps	B/W(Mbps)	CPU usage	MTU
Dell Precision 670 (single core)	Ubuntu 10.04.1 x64 (64bit)	DXT	12~13	105~110	75%	8900
		Uncompressed	23~24	730~750	35%	8900
Dell Precision T3400 (quad core)	Ubuntu 10.04.1 x64 (64bit)	DXT	30	260~270	85%	1450
		Uncompressed	13	440	8%	1450

GPU acceleration approach

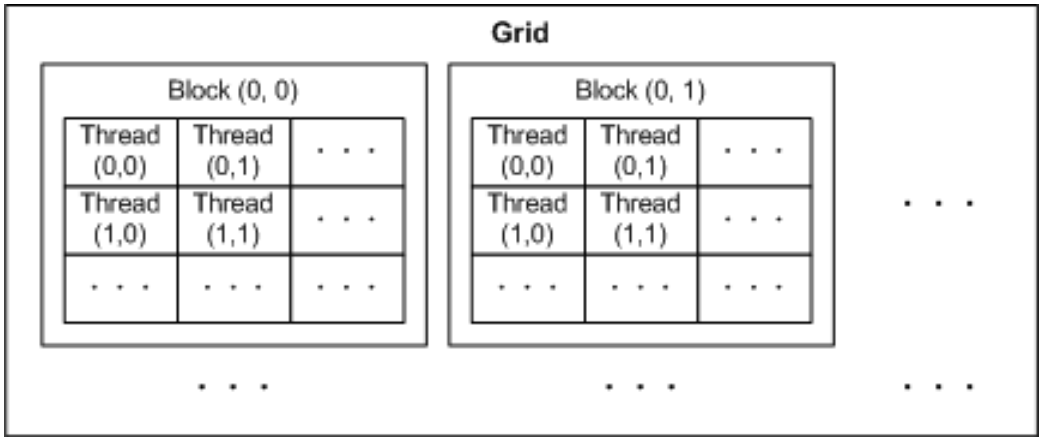
- OpenGL/Cg
  - high-level shading language developed by NVIDIA
  - suitable for GPU programming and it does not replace a general programming language
  - Cg compiler outputs DirectX or OpenGL shader programs
- CUDA: Compute Unified Device Architecture
  - parallel computing architecture developed by NVIDIA
  - gives developers access to the virtual instruction set and memory of the parallel computational elements in CUDA GPUs

CUDA Programming Model

- Parallel portions of an application are executed on device (GPU) as kernels
  - One kernel is executed at a time
  - Many threads execute each kernel
  - Kernels are lunched in grids



- Threads and Blocks have IDs
  - Each Thread can decide what data to work on



DXT Compression using CUDA

- Each pixel block is processed by CUDA Thread Block
  - Many pixel block compression simultaneously
  - Enables CUDA performance scalability
- CUDA Thread processing
  - Unpack color space each pixel
  - Compute DXT color index each pixel
- Apply FastDXT algorithm
  - Min/Max Color selection
  - Emit color indices

Reference

1) OpenGL DXT texture compression, http://www.opengl.org/registry/specs/EXT/texture_compression_s3tc.txt.
2) Libsquish, http://code.google.com/p/libsquish.
3) J.M.P. van Waveren "Real-Time DXT compression" May 20th 2006 © 2006, Id Software, Inc., http://www.intel.com/cd/ids/developer/asmo-na/eng/324337.htm.
4) L. Renambot, B. Jeong and J. Leigh, "Real-time compression for high-resolution content," Proceedings Access Grid Retreat 2007.
5) W. R. Mark, R. S. Glanville, K. Akeley and M. J. Kilgard, "Cg: A system for programming graphics hardware in a C-like language," ACM SIGGRAPH 2003
6) Nvidia-texture-tools, http://code.google.com/p/nvidia-texture-tools.
7) NVIDIA CUDA, http://developer.nvidia.com/object/cuda.html.
8) SAGE, http://www.sagecommons.org.
9) WIKIPEDIA, http://en.wikipedia.org/wiki/S3_Texture_Compression

Gwangju Institute of
Science & Technology

Thank you!

Send Inquiry to jongwon@nm.gist.ac.kr

http://nm.gist.ac.kr