Full GPU Image Processing Pipeline

for Camera Applications

Serzhenko Fyodor, Uvarov Andrey

Fastvideo, Russia

Key Points

- Implementation of full image processing pipeline on GPU for real time camera applications
- We can do that faster and with better quality in comparison with solutions on CPU, DSP, FPGA
- GPU Image Processing SDK from Fastvideo

Problem to Solve

Real time image processing

- Industrial cameras with high resolution and high frame rate
- Multiple camera systems
- Broadcasting solutions
- High speed cameras

What is common for all cases: data rates up to ~GPix/s

Image Processing in Realtime

Modern machine vision cameras

- Image sensor resolution from VGA to 30 Mpix
- Frame rate up to 500 fps
- Bit depth: 8 – 12 bits
- High speed interface to PC
- Multiple camera systems

Hardware and Software

- Image sensor with high data rate
- External interface to send data to PC
- Adapter or frame grabber to get data on PC side
- Software for real time data acquisition and image processing

Modern Image Sensors

Model	Resolution	Frames per second	Bits	Data rate
CMOSIS CMV20000	5120 x 3840	30 fps	8/12 bit	560 Mpix/s
CMOSIS CMV12000	4096 x 3072	100 fps	8/12 bit	1200 Mpix/s
Onsemi VITA 25K	5120 x 5120	53 fps	8/10 bit	1325 Mpix/s
Onsemi LUPA 3000	1696 x 1710	485 fps	8/10 bit	1340 Mpix/s
Alexima AM41	2368 x 1728	500+ fps	8/10 bit	> 2000 Mpix/s

External Camera Interfaces

	GigE	IEEE 1394b	USB-3.0	CameraLink	CoaXPress	PCIE x8 Gen2
Data rate (one cable)	1 Gb/s	0.8 Gb/s	6 Gb/s	2.5 Gb/s (Base)	3.125 Gb/s	8*5 Gb/s
Max data rate	1 Gb/s	0.8 Gb/s	6 Gb/s	6.8 Gb/s (Full) over 2 cables	25 Gb/s over 4 cables	40 Gb/s over 1 cable
Max cable length	100 m	5 m	100 m	10 m	70 m	7.5 m/~km
Cable type	Cat5	--	--	MDR26	Coaxial	Copper / Fiber

Near Future: PCI-Express Cameras

XIMEA xiB series with fiber optics interface

- Image sensor CMOSIS CMV20000 with resolution 20 Mpix
- Frame rate: 30 fps, Global Shutter
- Size: 60x60x38mm, Power: 4.5W
- Bit depth: 12 bits
- Real data throughput up to 1700 MB/s
- Synchronization via opto-isolated GPIO
- Canon EF-mount active lens interface
- Remote control of aperture, focus and zoom

Advantages of PCI-Express Cameras

- PCI-Express – modern industry standard
- One of the fastest interfaces on PC (high bandwidth)
- PCIE Hard IP from FPGA manufacturers
- PCIE IP Cores and drivers (Windows, Linux)
- Integration with fiber optics cables (long cable)
- No need in frame grabbers
- Camera has direct and fast access to CPU/GPU memory

Possible Solutions for Image Processing

- In-Camera HW processing on FPGA, DSP, CPU
- Image processing on frame grabber
- Image processing on CPU with SSE/AVX
- Image processing on GPU

Basic Image Processing Pipeline

- Image Acquisition
- Preprocessing
- Demosaicing
- Resize / OpenGL output
- JPEG compression

Extended Image Processing Pipeline

- GPU Direct option to get images on GPU from a camera
- Unpacking module for various raw formats
- Wavelet denoising
- Color correction
- Image tiling and batch processing
- Crop/Resize/Sharp and pyramid images
- OpenGL output
- MJPEG integration with FFMPEG

Preliminary Stages

- In-Camera HW image processing
- Camera-to-PC data transfer over cables
- Image acquisition on PC side
  - Data transfer to CPU memory
  - RDMA GPUDirect

RDMA for GPUDirect

Solution to get data at GPU with min latency

- Choose right GPU and 64-bit Linux OS
- Grabber and GPU should share the same root complex
- Driver modification (memory allocation on GPU)
- DMA data transfer from camera to GPU over PCI-Express

Image Preprocessing on GPU

- Image data unpacking and rearranging
- Fixed-Pattern Noise subtraction
- Vignetting
- Bad pixels removal
- White balance

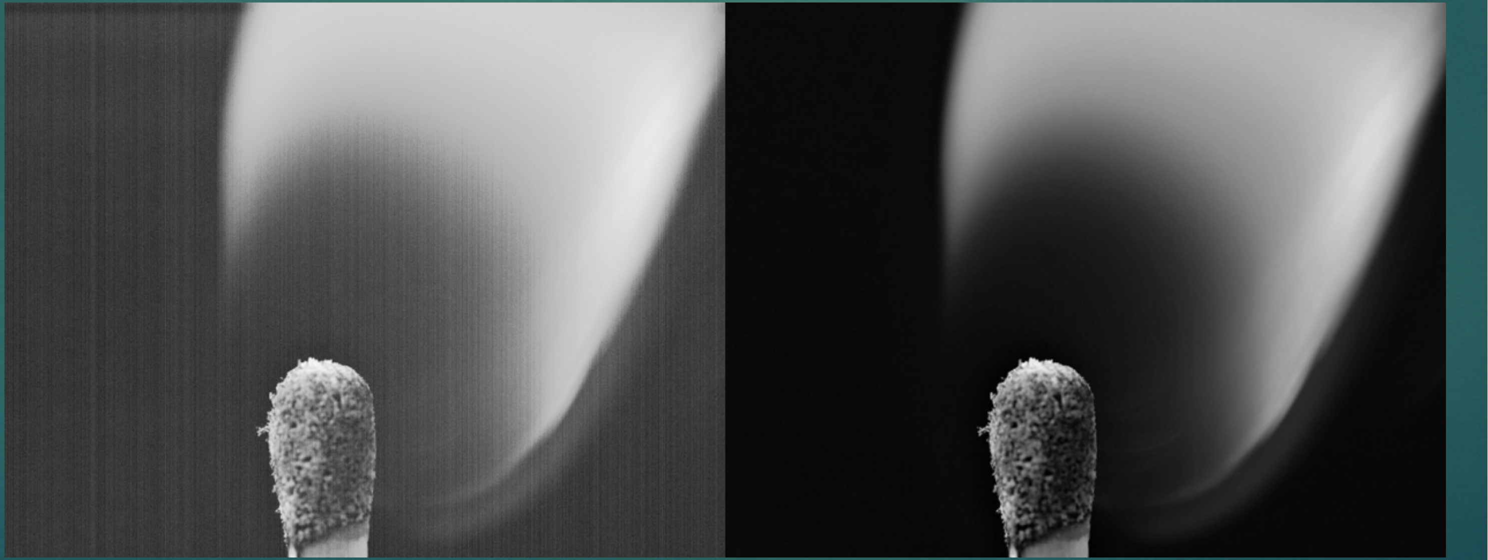
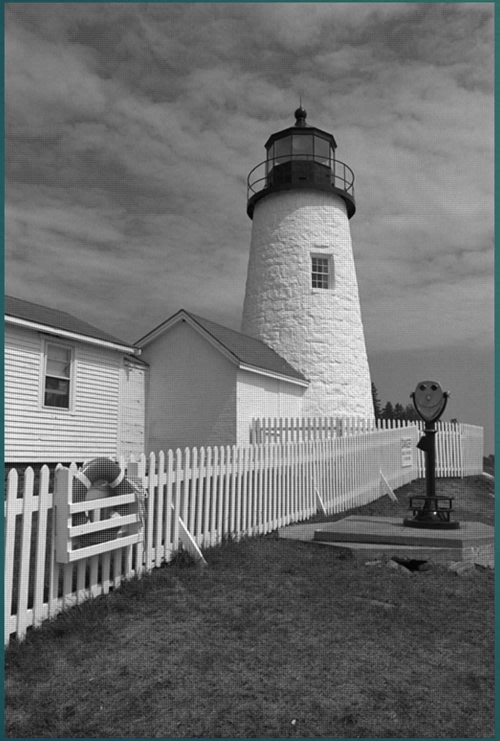
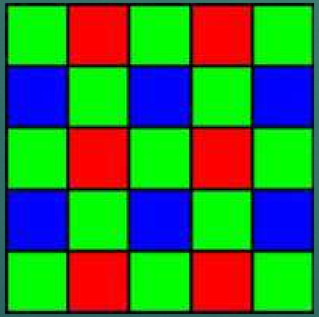


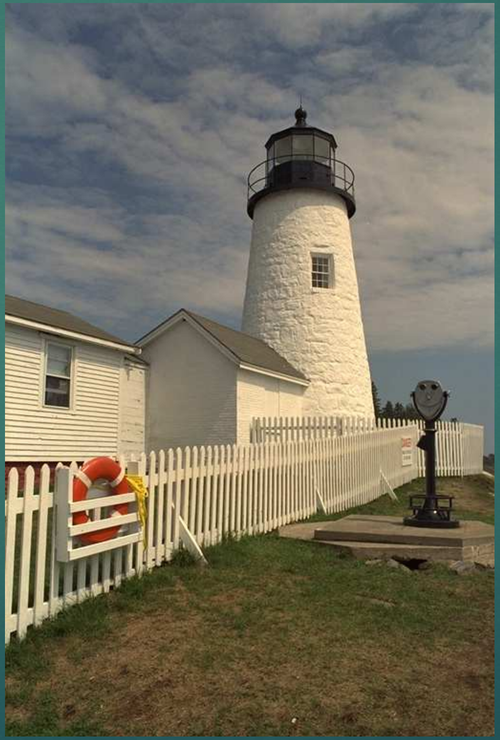
Image Demosaicing (Debayer)

Demosaicing means image color interpolation from Bayer image sensor

(input data 8 bits, output data 24 bits)







Ideas for high quality demosaicing

Fastvideo

- Raw Bayer data prefiltering
- Local gradient calculations
- Green plane interpolation and filtering
- R-G and B-G interpolation

Performance of Demosaic Algorithms

Fastvideo

- Bilinear – bilinear interpolation for each pixel
- HQLI – High Quality Linear Interpolation with 5x5 kernels
- AHD – Adaptive-Homogeneity-Directed algorithm
- DFPD – Directional Filtering with a Posteriori Decision

	Bilinear	HQLI	AHD	DFPD
Fastvideo on Titan GPU		2400 MB/s		1770 MB/s
IPP on CPU Core i7 3770	2800 MB/s		14 MB/s	
GPU kernel performance		21000 MB/s		6000 MB/s
SSIM	0.873	0.965	0.968	0.978
PSNR (dB)	30.4	36	37.4	39

Image Denoising on GPU

Fastvideo

Wavelet Denoising with Thresholding

- CDF discrete wavelet transform
- Hard or soft thresholding with edge preservation
- Inverse DWT

Image Tiling on GPU

- Real time image partitioning
- Quick method to retrieve ROI with min overhead
- No additional artifacts because JPEG is block-based
- Tiling means data rearrangement in GPU memory

Image Resizing

- Where we need resize: almost for any image on PC, laptop, projector, tablet PC, smartphone, TV, etc.
- Who need that: any user
- What we need: quality and performance
- Solution: image downsize prior to show it on monitor

Image Resize Quality

Fastvideo

- Bicubic (Photoshop)
- Wavelet CDF 9/7
- Fastvideo HQ
- Method: downsampling + upsampling
- Metrics: PSNR (dB)

Scaling ratio	1.1	1.3	1.5	1.7	2.0	3.0	4.0	6.0	8.0
Bicubic	52.4	50.4	49.1	48.1	46.8	43.4	40.6	36.8	34.7
Wavelet 9/7					47.7		41.3		35.4
Fastvideo HQ	56.6	52.6	51.0	49.9	48.0	44.8	41.9	38.1	35.7

OpenGL Interoperability

- All data is already in GPU memory
- All we need – just to show data on the screen
- No need in resize, image is ready
- Minimum latency for image output

CUDA JPEG Codec

In 2011 we have created the first fully parallel CUDA JPEG codec.

Main idea: JPEG algorithm consists of consecutive stages and it’s applied to blocks 8x8 or 16x16 pixels. Parallel block handling for all stages is a must for fast JPEG codec.

Key for fast decoding: presence of Restart Markers in bitstream.

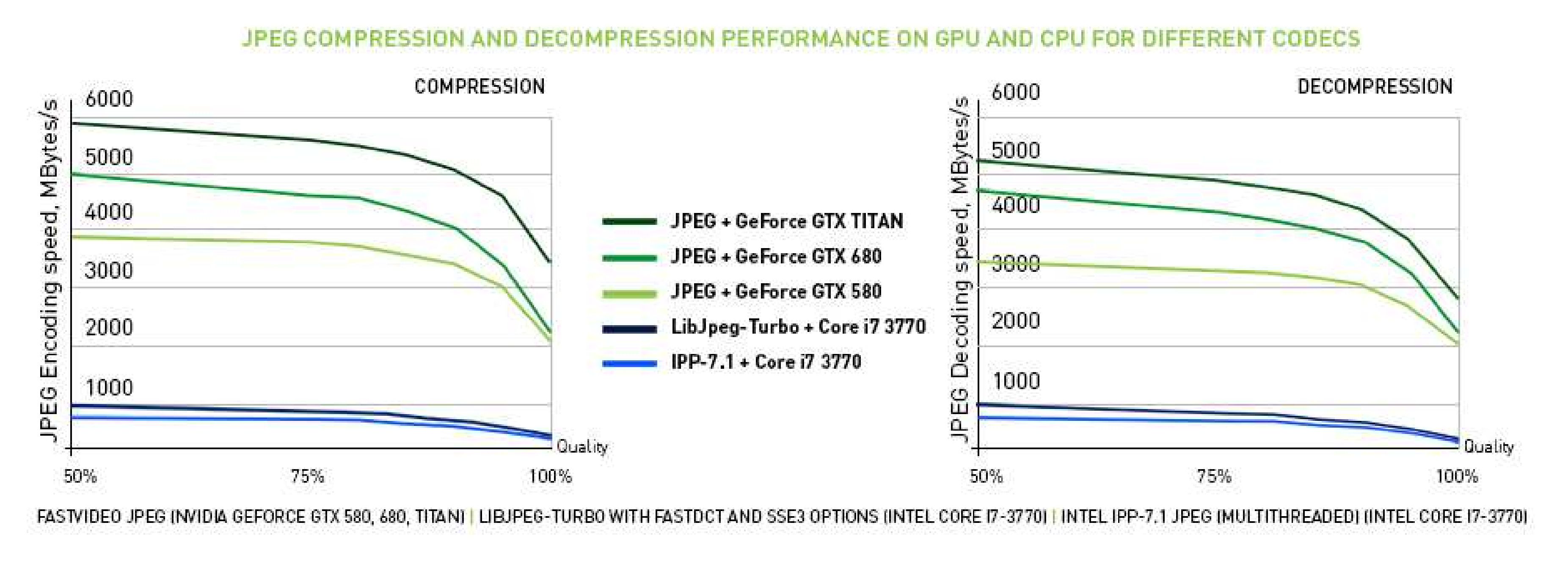
Features of CUDA JPEG Codec

- Compliance with JPEG Baseline Standard
- JPEG quality from 1 to 100%
- Subsampling modes 4:2:0; 4:2:2; 4:4:4
- JPEG Encoding 2500 – 5500 MB/s on GeForce GTX Titan
- JPEG Decoding 2000 – 5000 MB/s on GeForce GTX Titan

Fastvideo

CUDA JPEG results for 580/680/Titan

- Host-to-Device and Device-to-Host transfers are included
- Win-7 (32-bit), NVIDIA drivers 332.21



JPEG COMPRESSION AND DECOMPRESSION PERFORMANCE ON GPU AND CPU FOR DIFFERENT CODECS

FASTVIDEO JPEG (NVIDIA GEFORCE GTX 580, 680, TITAN) | LIBJPEG-TURBO WITH FASTDCT AND SSE3 OPTIONS (INTEL CORE I7-3770) | INTEL IPP-7.1 JPEG (MULTITHREADED) (INTEL CORE I7-3770)

JPEG and Image Distortion

- JPEG is not always guilty for insufficient image quality
- There are other lossy (irreversible) stages in the pipeline:
  - Preprocessing
  - Demosaicing
  - Denoising
  - LUT
  - Resize / Sharp

Solution: use “visually lossless” JPEG quality 85-100% and always check quality losses for other stages of the pipeline

Combined Demosaic + JPEG

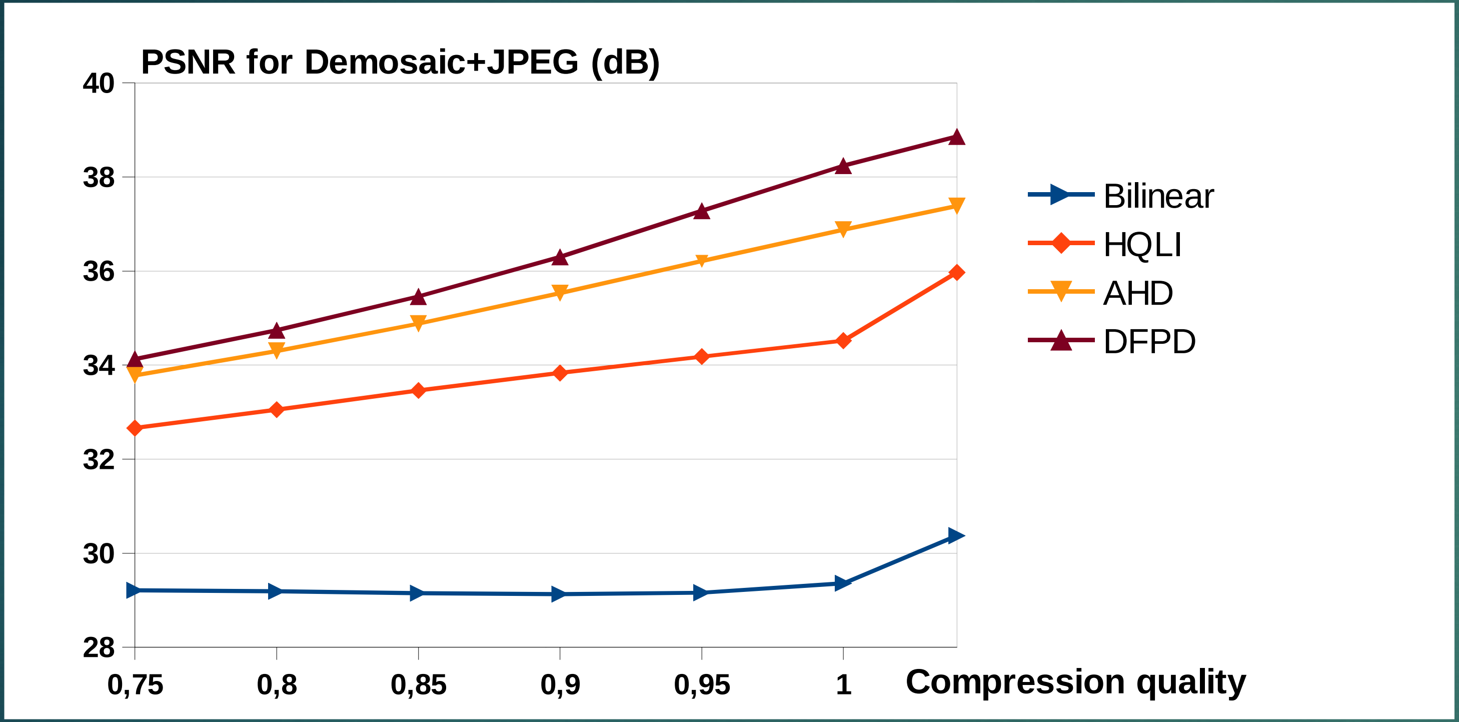
Fastvideo

- Tripled data size after demosaicing
- Transfer time over PCI-Express bus is important
- Combined solution saves time for CPU-GPU transfers
- Low-quality demosaic is unacceptable

Performance benchmarks for Demosaic and JPEG Encoder

	1920 x 1080	2048 x 2048	4096 x 4096
Demosaic (DFPD)	1.6 ms	3.0 ms	10.2 ms
JPEG (90%, 4:2:0)	2.3 ms	3.2 ms	10.5 ms
Demosaic + JPEG	2.6 ms	3.7 ms	11.7 ms

Image Quality for Demosaic + JPEG



Final Benchmark on GPU (Titan)

- CMOSIS image sensor CMV20000, 5120x3840, 12-bit, 30 fps
- GPU GeForce GTX Titan
- Host to device transfer ~1.5 ms
- Demosaic ~3.1 ms
- JPEG encoding (90%, 4:2:0) ~7.8 ms
- Device to Host transfer ~1.3 ms
- Total: ~13.7 ms

P.S. This is the benchmark for PCIE camera xiB-20 from XIMEA

Batch Image Processing on GPU

We need batch mode to get maximum GPU performance

- Preprocessing
- Demosaicing
- Denoising
- Tiling
- Resizing
- JPEG encoding

Integration with FFMPEG

- Compatibility with various video containers
- MJPEG output format for camera stream
- Interoperability with other applications

CUDA Optimization Techniques

- Concatenation of simple CUDA kernels to reduce Global Memory I/O (all stages of DFPD demosaic algorithm are integrated in one kernel)
- Register cache to reduce Shared Memory I/O
- Two code paths: one for central area and another for the edges
- Careful resource management (registers, shared memory) for optimal SMX utilization
- Reduce communications between Host and Device by moving all processing to GPU

Camera Applications

- Machine vision and computer vision
- XIMEA solutions – www.ximea.com
- Imperx solutions – www.imperx.com
- Multi-camera image processing on single GPU
- Real time image processing for broadcasting
- Video conferencing
- High speed imaging

XIMEA Applications

- Flat panel display control
- Printed circuit board examination
- Persistent stadium/border security, wide area surveillance
- City and aerial mapping, UAV
- High definition scanner

IMPERX Applications

29 MP@5FPS CCD Camera built on TRUESENSE image sensor

- Aerial imaging
- Medical and scientific imaging
- Automation and Inspection

Bayer Demosaicing performance

- 91 ms for Intel IPP Fast algorithm (Intel® Core™ i7-3770 @ 3.4 GHz)
- 33 ms for DFPD high-quality algorithm (GeForce GTX 650 Ti, 2 GB)

Non-Camera Applications

Fastvideo

- Image processing speedup to image viewers, photo editors
- High performance batch image processing
- Fast resize and thumbnail generation
- Cloud solutions for image processing
- Medical imaging

Standard Task for Photo Hosting

- Get HTTP query from remote user
- Find corresponding image in database
- Load full-size JPEG image from database to CPU memory
- JPEG decoding
- Image resizing according to resolution of user’s device
- JPEG encoding
- Store image to HDD/SSD or send it to remote user

Benchmark for Photo Hosting Task

Task description: Load-Decode-Resize-Encode-Store

- Image load ~1.5 ms for 2048x2048 jpg image
- JPEG decoding ~3.4 ms
- Downsize to 1024x1024 with bicubic algorithm ~0.7 ms
- JPEG encoding (quality 90%, 4:4:4) ~3.4 ms
- Image store ~1.0 ms
- GPU processing time ~7.5 ms (133 images per second)
- Total time ~10 ms

What's New in Photo Hosting Solution

- This is GPU-based solution
- Very high performance
- Opportunity to use better sharp and resize algorithms
- Improved accuracy for the whole pipeline (float DCT in/out)
- Excellent scalability

This sample project is included in Fastvideo Demo SDK

What we could offer to our customers

Fastvideo

- Standard GPU Image Processing SDK
  - Preprocessing functions
  - Demosaic
  - LUT
  - Resize + OpenGL
  - JPEG Encoder and Decoder
- Custom SDK for particular task
- Algorithm design
- Software optimization

Our Tasks for Near Future

- High quality resize (downsampling and upsampling)
- 12-bit JPEG
- Batch mode optimization for JPEG codec
- Design of better 16-bit Demosaic

Long Term Strategy

Fastvideo

- More image processing algorithms in our SDK
- More complicated parallel algorithms with higher quality
- JPEG2000 and more lightweight custom codecs
- Code optimization, use of the latest GPUs and APIs
- Parallel algorithm design and implementation

The Most Important Results

- High speed and high quality Demosaic on GPU
- Super fast CUDA JPEG Codec
- SDK for full image processing pipeline on GPU

Links & Contacts

- Fastvideo site for GPU image processing www.fastcompression.com
- Contacts: info@fastcompression.com

Camera manufacturers with GPU image processing software

- XIMEA GmbH – www.ximea.com
- Imperx Inc. – www.imperx.com