



High-speed image acquisition with real-time GPU processing

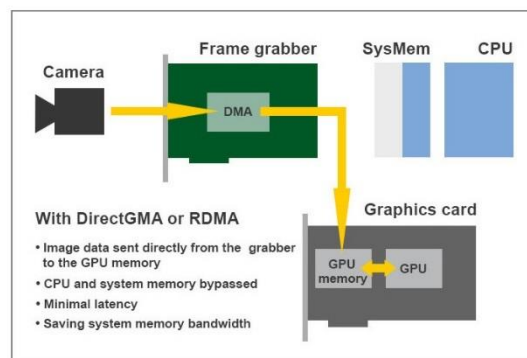
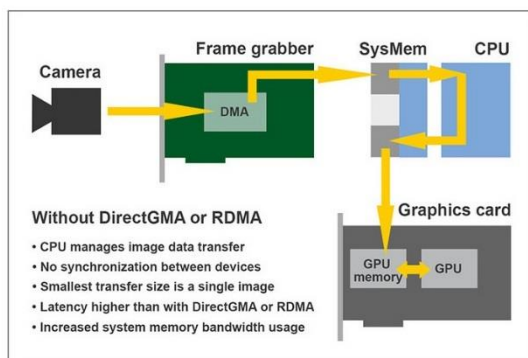
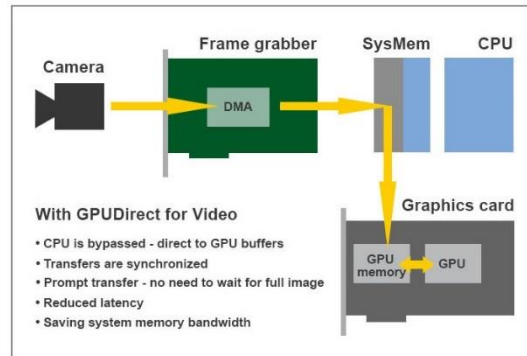
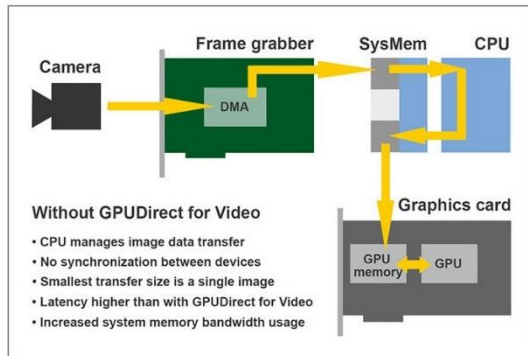
Alongside the adoption of FPGAs, it's interesting to take a look at how image processing via a GPU remains an immensely efficient and flexible option. Offering lower implementation costs and accessibility, a GPU alleviates the burden on a CPU, freeing up memory and capacity for other functions

High-speed imaging via a GPU can add value to a vision system

Modern GPUs are extremely efficient at processing images and graphics. Their structure makes them particularly well suited to applications where large blocks of data need to be processed in parallel. As cameras and sensors become more powerful, more and more data needs to be processed. In the machine vision market, an increasing number of industries are moving towards 100% in-line inspection and, as a result, imaging tasks are getting more problematic. A greater variety of different and more complex objects needs

to be scanned in a shorter period of time using multiple sensors.

Due to the higher quality standards required, heavy computational algorithms are having to be used, and of course the introduction of Deep Learning in the machine vision market is adding to this. When using GPUs, all levels of processing can be carried out without the need for a CPU. This has led to developments such as IBM unveiling a new benchmark in 2017; they were able to train a logistic regression classifier 46 times faster than the previous TensorFlow record (90 seconds compared to 70 minutes). This relied on distributed



training, GPU acceleration and new algorithms.

The two leading suppliers of GPU technology are NVIDIA and AMD, who provide the APIs Direct for Video and DirectGMA respectively, which can both be used in conjunction with acquisition cards. With these APIs, the system buffers and the CPU can both be skipped as the GPU memory is made directly accessible to the frame grabber, which can be used to enable high-speed, real-time image acquisition with very low latencies.

FPGAs also provide technology for parallel computing, and many manufacturers, including ourselves, use both FPGA and GPU technology in their hardware. In general, GPUs have several advantages over FPGAs, starting with lower implementation costs. Furthermore, software development is so much quicker and more straightforward than developing FPGA firmware, and there are more engineers available in the market with the necessary skills and knowledge. Additionally, it is easier to use third-party software in the vision system.

Applications suited to GPU processing

Most GPUs have built-in video codecs that can be used to accelerate video compression, so applications needing to encode video streams are particularly well suited to GPU processing.

Frame grabbers and GPUs are an ideal combination for applications where high-data-rate, real-time image processing on the fly is required. One of our FireBird frame grabbers, for example, is currently being used in a novel system developed by Norwegian company, Piql. Piql offers secure, searchable, high-volume digital data storage on 35mm film. Their unique application digitizes data from various mediums, applies OCR and indexing (for fast searching), then digitally prints the data onto highly-durable 35mm film using coding technology similar to QR codes. The data may then be accessed using their specialist scanner to read the data back from the film. All with the result at the end that the data can be stored securely for at least 500 years. Piql chose an Active Silicon Firebird frame grabber to install in their readers due to its high-speed image acquisition capabilities, cost-effective

pricing and its off-the-shelf compatibility with other system components. The Piql storage process writes data at 20-24 frames per second onto 150mm of 35mm film. Each frame generates 80 MB of data. Therefore 1.6 GB of data are processed every second, continuously. This massive amount of data and high compute power could not be processed on a CPU, so Piql investigated the options of both GPUs and FPGAs. In the end they opted for an NVIDIA GPU due to its superior system flexibility. More options were available in terms of GPU cards, configurations, frame grabbers and other compatible components.



Another interesting case study involves medical imaging. Active Silicon currently produces a USB 3.0 Embedded Vision System for a specific medical application in the field of computer vision assisted surgery. Our current system processes the image data of up to three USB 3.0 cameras, two visible light cameras for stereo vision and one non-visible light camera, to create a fluorescent image and for image fusion. We are now developing the next generation system which will acquire, process, display and record images in 4K. Images from up to four 4K 60 Hz cameras will be used, generating a massive amount of data. FPGAs, ASICs and GPUs were considered as solutions to this massive data processing requirement – a CPU simply cannot offer enough processing power. While FPGAs and GPUs offered identical performance in processing under test conditions, a GPU was seen as the best solution because the

software development is so much quicker and more straightforward than developing FPGA firmware. ASICs offered no apparent benefits. Software already compliant with medical ISO (13485) can be purchased and installed on a GPU, eliminating the need for further compliance testing. An NVIDIA GPU was selected over an AMD one; the product is well developed, has considerable libraries and support available, and more engineers are familiar with NVIDIA CUDA coding, offering simpler recruitment options.

GPU Processing and Deep Learning

GPUs and Deep Learning are used, for example, in autonomous vehicles. Automation is divided into four areas – sensing, perception, prediction and planning. Sensors being used are radar, LiDAR and cameras. Deep learning is used for road sign recognition, pedestrian detection and lane monitoring. Images are complex, massively diverse and affected by conditions including lighting and weather. It is compulsory that processing and resultant actions need to be taken in real-time. Because of their ability to manage huge amounts of data, GPUs are currently the best solution to this image processing. However, challenges include managing the energy consumption levels and the temperature environment associated with vehicles. One solution to the energy problem is to train neural networks on GPU servers external to the vehicle and then transfer data to on-board units once completed. NVIDIA and Intel are both focusing on producing GPUs specifically for autonomous driving: NVIDIA Tesla V100 achieves a computing speed of 120 Tera Operations Per Second with power consumption of 300W. The NVIDIA



Jetson TX1 mobile GPU offers a speed of 300 Giga Operations Per Second on a VGG-16 and a peak performance of 1 Tera Operations Per Second with only a 10W draw.

Bringing Deep Learning to your system – what to look for

Deep Learning in conjunction with machine vision is an exciting and fast-moving tool for inspection and the massively parallel architecture of a GPU makes it much more suitable than a CPU. Deep Learning relies heavily on matrix multiplications, these algorithms simply cannot be effectively handled on a CPU as machines running Deep Learning are able to find patterns in data or features on their own and only limited training is needed. So, what should you look for when bringing Deep Learning to your system? The important features of a GPU running deep learning are a high memory bandwidth, high processing power and a large video RAM size. NVIDIA has been leading the way and is therefore a thoroughly tried and tested option. CUDA is a well-established language compatible with all major deep learning frameworks. NVIDIA Jetson supports Artificial Intelligence capabilities at the edge for portable medical devices, robotics and smart cameras. The Jetson TX2 contains a NVIDIA Pascal GPU with 256 CUDA cores. The second biggest player in the market, AMD, offers more cost-effective products and some of the AMD GPUs have also increased performance and increased VRAM size.

Implementing high-speed image acquisition

Processing via GPUs is a simple and cost-effective way to speed up image acquisition which can be implemented easily and quickly into most modern vision systems. At Active Silicon, we are leaders in providing reliable and robust components which are compatible with GPU technology. Our Camera Link frame grabbers achieve clock rates of to 85 MHz while our new CoaXPress v2.0 quad boards offer data at a transfer rate of up to 50 Gbps. Fitting easily into existing systems and requiring readily-available software skills, speeding up imaging and implementing Deep Learning has never been easier.



Author | Frans Vermeulen, Head of Sales and Marketing, Active Silicon