GPU Adoption in Intelligent Things - A Public Safety

Abstract

Autonomous systems such as driverless cars, drones, and robots are moving from the realm of science fiction to reality, fueling an insatiable demand for accelerated, real-time application processing. Simultaneously, emerging technologies such as augmented and virtual reality (AR and VR), enhanced graphics engines, and conversational platforms, require intense processing power beyond that provided by standard CPUs.

Graphics Processing Unit (GPU), a supercharged form of computing, is the answer to the evolving trends. While the GPU was originally conceived to support advanced graphics for gaming and design, the form-factor and unused processing power of GPUs is beckoning technology pioneers in other areas. Businesses are evaluating the ability of GPUs to enable high-performance tasks such as machine or deep learning, video analytics, AI-driven intuitive decision making, and gamification.

This paper focuses on leveraging on-board GPU to execute advanced public safety use-cases in unmanned aerial vehicles (UAV) and autonomous vehicle ecosystems, and highlights two use-cases: autonomous vehicles and Beyond Visual Line of Sight (BVLOS) drones.

GPU: Modern Day CPU for the Age of AI

The success of emerging autonomous systems such as drones and self-driving cars depends on their navigation systems' ability to quantify uncertainty through real-time simulation, which calls for rapid processing. This is where GPU comes in.

The GPU provides superior processing power, memory, bandwidth, and efficiency, over a traditional CPU. It is 50 to 100 times faster in processing tasks that require multiple parallel processes, such as machine learning algorithms and Big Data analytics.[1] Unsurprisingly, this is leading to growing usage of GPU in scenarios that require high performance computing (HPC) such as gaming engines, video analytics, and machines, including autonomous cars, humanoids or robots, and UAVs. Thanks to its incredible compute performance, calculation density, time savings, and overall system speed, the modern GPU has morphed from a graphics processor into a general purpose co-processor that sits alongside the CPU in today's advanced data processing platforms.

Leveraging On-board GPU for Accelerated, Real-time Processing

The rise of GPU, however, does not signal the death of the CPU. In fact, the combined power of GPU and CPU processing is essential to significantly accelerate analytics, engineering, consumer, and enterprise applications. For instance, during the performance test of a face detection system we conducted, it was found that using a combination of GPU and CPU cut the processing time in half as compared to CPU processing alone (see Table 1).

	GPU	CPU		
	Real data (2.63 MB per frame	Extrapolated (200 GB per frame	Real data (2.63 MB per frame	Extrapolated (200 GB per frame
GPU processing time	90.828 milli-seconds (ms)	1.94 hour		
Host to device transfer	81.93 micro-seconds (us)	6.38454 second	GPU is not used	GPU is not used
Host to device transfer	43.89 us	3.4183		
CPU time to execute complete time	750 ms	16 hour	1400 ms	32 Hours

Table 1: Processing time for GPU Vs CPU for a face detection system

So how can autonomous systems leverage on-board GPU for accelerated processing and high performance? The need for parallel programming across heterogeneous systems has evolved programming languages such as OpenCL and CUDA. These enable customization of advanced applications (video or image analytics leveraging deep learning models) for deployment on GPU. The CPU (on standard x86 systems or equivalent) then transfers the processing to the on-board GPU.

GPU processing scores over CPU because of the following four parameters:

- Compressed GPU execution (time taken by GPU to execute an algorithm)
- CPU-GPU transfers (time taken for handovers)
- CPU wait time (time take for CPU to wait for GPU processing completion)
- Abbreviated CPU execution (time taken by CPU to execute pipeline)

What Does GPU Computing Bring to Autonomous Systems?

Autonomous systems augment navigation with high-quality sensor data to enable better interpretation of the surrounding environment, leading to improved accuracy and performance. However, increasing the amount of input data alone does not drive enhanced accuracy. The system must be equipped to process the data quickly enough to make a difference in its real-time response and capabilities. Given its parallel, throughput-oriented nature, the GPU is a natural and excellent fit for solving the real-time processing challenges of autonomous systems. Here are two use cases:

Autonomous vehicle driving safety

Autonomous vehicles have the potential to change more than just how we get from one place to another. They can significantly reduce the number of accidents, cut fuel costs, accelerate speed, and even provide mobility to those who can't drive. Fueled by AI, embedded GPUs can make autonomous vehicles smarter by enabling features such as lane detection, lane departure warning, distracted driver warning, and road signs or pedestrian detection. An on-board GPU can be integrated with the dashboard camera for image acquisition of the autonomous vehicle front view, which can be further fine-tuned using the computer vision algorithm to track the latitude and longitude precisely.

Anticipating every possible scenario that a self-driving car might encounter is nearly impossible. Deep learning, which relies on powerful GPUs, helps tackle this challenge by analyzing humongous amounts of data to enable deep neural networks to learn, adapt, and improve their problem solving capabilities in real time. Geo-tagging of images and landmarks can help deep learning algorithms to dynamically select travel routes in autonomous vehicles, with sensor integration enabling reliable and precise on-road navigation.

GPU based advanced driver-assistance systems (ADAS) can also reduce human error, which accounts for 94% of car accidents in the US alone and 1.2 million deaths worldwide every year, according to data from the National Highway Traffic Safety Administration and the World Health Organization respectively.[2] Human error typically occurs when drivers are fatigued (frequently driving for 3-30 seconds with their eyes closed) and/or distracted (text messaging, reaching for an object at the back, or trying to operate the radio). Toyota is exploring the use of GPU-powered simulation and deep learning in making its self-driving cars smarter.[3] In the future, GPUs could also enable superior traffic control and surveillance through automatic number plate recognition systems.

Autonomous drone search and rescue operations

New use cases of drones extend far beyond delivering shipments. GPU computing can enable improved spatial awareness when operating drones beyond line-of-sight as it impacts how drones communicate with each other and with other unmanned aerial systems (UAS). Coordinated communication between the drone and the flight management system enables real-time flight monitoring for improved flight path conformance and dynamic adjustments to drone operations. Seamless drone traffic management helps improve flight efficiency and reduces air collisions, paving the way for automated search and rescue operations, and quicker and safer disaster recovery.

Intel has partnered with NASA, the FAA, and industry partners to introduce its Aero Ready to Fly Drones powered by GPU computing, that are designed to conduct advanced mission critical safety operations.[4] BVLOS drones can also be used in case of forest fires and natural disasters such as earthquakes and storms, to track, trace, and rescue humans. Researchers from the DalleMolle Institute for Artificial Intelligence used deep learning to train a drone to navigate a previously-unseen trail in a densely wooded forest. The drone was able to determine the correct direction of the trails with 85% accuracy, while humans could achieve only 82% accuracy.[5]

Unleashing the Power of GPU Computing

By simulating human intelligence powering systems, GPU computing helps unlock the mysteries in humungous amounts of data, in real time, to augment autonomous systems. In addition to the use cases discussed, GPU computing can enhance enterprise data center security by deploying an anti-virus application on GPU to enable email attachment scanning. Similarly, its application in virtual reality is on the rise, wherein participants can test and apply their safety skills in various simulated scenarios without being in actual danger.

Supply chain and inventory control is another key area where GPU computing can enable real time data analytics visualization for shipment and in-store inventory optimization. Walmart's Data Cafe powered by GPUs analyzes 200 streams of internal and external data, including 40 petabytes of recent transactional data to make faster sales and inventory decisions.[6]

As GPU applications gain rapid traction, understanding its capabilities and evaluating the possibilities, can help organizations outdo competition in today's uber competitive markets.

References

1. Stream HPC, GPUs and Gartner's Top 10 Strategic Technology Trends For 2017, March 2017 (accessed May 2018), https://streamhpc.com/blog/2017-03-27/gpus-gartners-top-10-strategic-technology-trends-2017/?upm_export=print
2. Nvidia, 2016 Sustainability Report, Deep Learning, Autonomous Cars, Sep 2016, Accessed May 2018, http://www.nvidia.com/object/autonomous-cars.html
3. Nvidia, Toyota Exec Explains Why Simulation Key to Autonomous Driving, April 2016, Accessed May 2018, https://blogs.nvidia.com/blog/2016/04/07/toyota-gtc/
4. Intel, Newsroom, Bird's-Eye View: How Traffic Management Will Lead to Safer Skies, June 2017, Accessed May 2018), https://newsroom.intel.com/editorials/birds-eye-view-how-traffic-management-will-lead-safer-skies/
5. Nvidia, News Center, Autonomous Search-and-Rescue Drones Outperform Humans at Navigating Forest Trails, Feb 2016, Accessed May 2018, https://news.developer.nvidia.com/autonomous-search-and-rescue-drones-outperform-humans-at-navigating-forest-trails/
6. RT Insights, Walmart's Gigantic Private Cloud for Real-Time Inventory Control, Jan 2017, Accessed May 2018, https://www.rtinsights.com/walmart-cloud-inventory-management-real-time-data/

About The Authors

Partha Datta

Partha Datta is a senior domain consultant leading the NextGen R&D group of TCS' Technology business unit. He has over 19 years of experience in the telecom industry helping clients conceptualize technology solutions for business process transformation in the areas of software defined networking in a wide area network (SD-WAN), drones, robotics, and AI. Datta holds a Master's degree in Instrumentation from the Indian Institute of Science (IISc), Bangalore, India.

Saloni Jain

Saloni Jain is a lead solutions architect with TCS' Technology business unit. She has over eight years of experience in the areas of drones, robotics, and GPU, and is an avid contributor to the open-source community. Jain specializes in implementing proofs-of-concept (PoCs) and solution enablers by leveraging the drone ecosystem.

Deepak Jain

Deepak Jain is a lead solutions architect with TCS' Technology business unit. He focuses on high performance computing (HPC) and real-time applications. Jain specializes in the implementation of PoCs leveraging GPU to accelerate the performance of real-time ecosystems pertaining to image or video analytics and equivalent cognitive solutions.

Cheshta Kumar

Cheshta Kumar is part of the NextGen R&D group of TCS' Technology business unit. She focuses on developing GPU or HPC related solution accelerators to enable differentiated services in the computing (image or video analytics) ecosystem. With more than three years of experience, Kumar is responsible for implementing PoCs and use cases to address industry challenges.

Contact

Visit the Communications, Media & Technology page on www.tcs.com

Email: marketing.techbu@tcs.com

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w
Feedburner: http://feeds2.feedburner.com/tcswhitepapers

About Tata Consultancy Services Ltd (TCS)

Tata Consultancy Services is an IT services, consulting and business solutions organization that delivers real results to global business, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled, infrastructure, engineering and assurance services. This is delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development. A part of the Tata Group, India's largest industrial conglomerate, TCS has a global footprint and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

TCS Design Services I M I 08 I 18