Researcher Day, School of Engineering, Trinity College Dublin, The University of Dublin, Ireland, 25th April 2024

EMERALD: ENERGY EFFICIENCY IN AI TOOLS FOR POST PRODUCTION WORKFLOWS

V, Hareesh.[1], Shanker, Shreejith.[2]
[1,2] Reconfigurable Computing Systems Lab, Department of Electronic and Electrical Engineering, Trinity College Dublin, The University of Dublin
[1] veekanch@tcd.ie, [2] shankers@tcd.ie

INTRODUCTION

Video dimensionality has increased recently, offering greater temporal and spatial resolutions together with a wider color gamut for high dynamic range (HDR) visuals [1]. High-end production and post-production tasks generate and process massive data volumes, of the order of many petabytes per day, requiring extremely high compute capability achieved through clusters of processing hardware including CPUs, GPUs and specialised digital processing (ASICs/FPGAs). Additionally, digital media accounts for over 80% of the internet traffic covering a spectrum of both entertainment and work-related applications [2]. Various signal processing accelerators and AI algorithms are actively being developed and utilised to assist artists in post-production tasks and to improve streaming efficiency, among other applications. However, the massive amount of data and acceleration requirements lead to significant energy utilisation, an issue that the media industry is becoming increasingly sensitive to.

METHODS

The project involves collaborations with partners across Europe, including the BBC, DISGUISE, FILMLIGHT, BRAINSTORM and MOG, and the University of Pompeu Fabra (UPF, Spain), with the aim to develop energy-efficient techniques and tools for tasks within the media processing and media delivery pipeline.

In the Horizon EU project, EMERALD, we aim to develop tools and techniques to provide insights into where and how the energy is consumed within the media pipeline through granular energy profiling, in conjunction with TCD contributions on optimising AI algorithms and efficient compression techniques to show the energy benefits that can be achieved in this domain. Power analysis is done using the using the tools such as NVIDIA System Management Interface (nvidia-smi) and INTEL running average power limit (RAPL) interfaces. The Intel RAPL interfaces facilitate the reporting of Intel CPU energy usage overall over several power domains [3]. The nvidia-smi is a command-line utility built on top of the NVIDIA Management Library (NVML) that helps manage and monitor NVIDIA GPU devices [4].

RESULTS

Software packages are used to analyse the power consumption data for the CPU and GPU while executing various tasks in Nuke 15.0v4 VFX and Film editing software. The results indicate that tools related to video processing consumes a significant amount of energy.

DISCUSSION

Numerous research publications have demonstrated that on-chip power sensors suffer from a large error rate and poor sampling intervals (usually in the range of tens of milliseconds) [5]. Consequently, relying on data from these sensors for optimizing algorithms is unreliable. Optimizing energy in tools for post-production workflows requires precise energy profiling and in-depth energy analysis.

REFERENCES

[1] *Vibhoothi, A. Katsenou, F. Pitié, K. Domijan and A. Kokaram, "Subjective Assessment of the Impact of a Content Adaptive Optimiser for Compressing 4K HDR Content with AV1," 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 2023, pp. 2610-2614.*

[2] *Cisco Visual Networking Index (VNI) Complete IP Traffic & Service Adoption Forecasts Global Update, 2014–2019. https://www.cisco.com/c/dam/m/en_us/service-provider/ciscoknowledgenetwork/files/527_2015_VNI_Complete_Forecast__Master-Global__CKN_6102015.pdf*

[3] *INTEL Running Average Power Limit Energy Reporting. https://www.intel.com/content/www/us/en/developer/articles/technical/software-security-guidance/advisory-guidance/running-average-power-limit-energy-reporting.html*

[4] *System Management Interface SMI https://developer.nvidia.com/nvidia-system-management-interface*

[5] *Yang Z, Adamek K, Armour W. Part-time Power Measurements: nvidia-smi's Lack of Attention. arXiv preprint arXiv:2312.02741. 2023 Dec 5.*