# Computer Vision on Tegra K1

Chen Sagiv
SagivTech Ltd.

# SagivTech Snapshot

- Established in 2009 and headquartered in Israel
- Core domain expertise: GPU Computing and Computer Vision
- What we do:
  - Technology
  - Solutions
  - Projects
  - EU Research
  - Training
- GPU expertise:
  - Hard core optimizations
  - Efficient streaming for single or multiple GPU systems
  - Mobile GPUs

SAGIVTECH

# Mobile Revolution is happening now !

- In 1984, this was cutting-edge science fiction in The Terminator

- 30 years later, science fiction is becoming a reality!

SAGIVTECH

# The vision

Computer vision

Machine Learning

Deep Learning

to run on Mobile GPUs

SAGIVTECH

# 3D Imaging is happening now !

- Acquisition –  Depth Sensors
- Processing – modeling, segmentation, recognition, tracking
- Visualization – Digital Holography

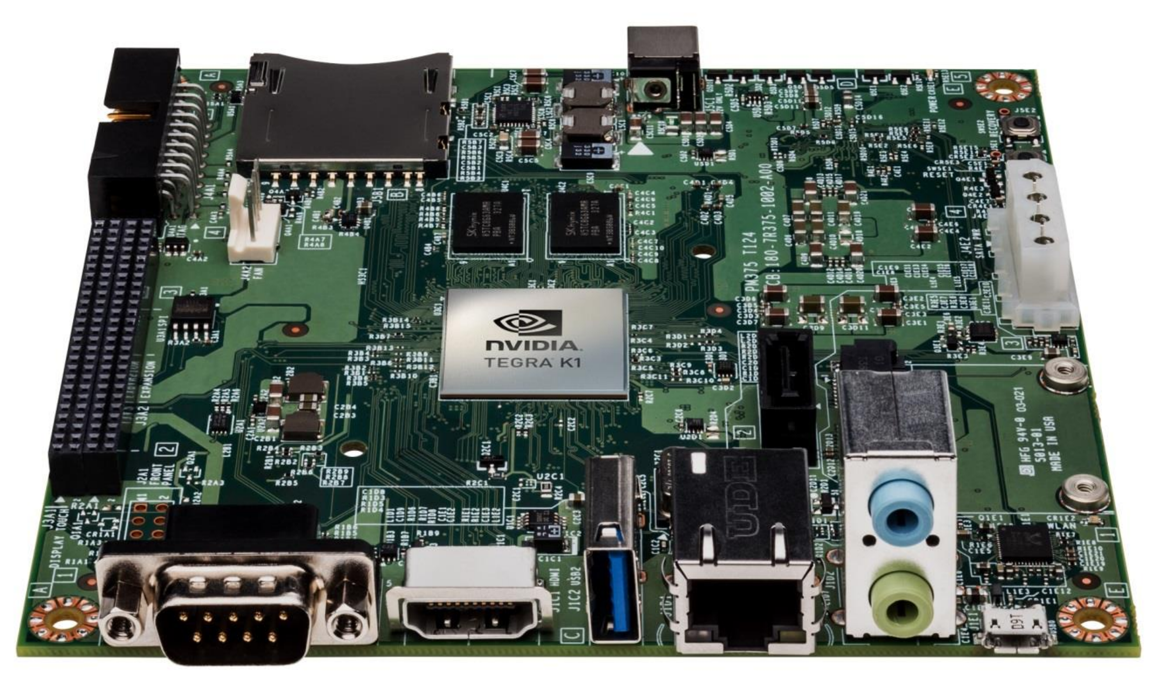
SAGIVTECH

# First Depth Sensing Module for Mobile Devices – on Tegra K1

- The Mission: Running a depth sensing technology on a mobile platform
- The Challenge: First time on Tegra K1
- Extreme optimizations on a CPU-GPU platform to allow the device to handle other tasks in parallel
- The Expertise:
  - Mantis Vision – the algorithms
  - NVIDIA – the Tegra K1 platform
  - SagivTech – the GPU computing expertise
- The bottom line: Depth sensing in running in real time in parallel to other compute intensive applications !

SAGIVTECH

# Tegra K1 & Jetson




SAGIVTECH

# Computer Vision on Tegra K1

- You can rely on the CUDA Eco System

- Having building blocks, e.g. Features Descriptors, for various  vision tasks

SAGIVTECH

# Bilateral Filter Acceleration on Tegra K1

| Image Size | 1 CPU Thread | 4 CPU Threads | GPU | Speedup |
|---|---|---|---|---|
| 256 x 256 | 630ms | 170ms | 2.8ms | x60 |
| 512 x 512 | 2550ms | 690ms | 12ms | x57 |
| 1024 x 1024 | 10300ms | 2720ms | 45ms | x60 |

SAGIVTECH

# Migrating from Discrete Kepler to K1

- In one word: Easy!
- Took only a few hours to transfer all the code.

SAGIVTECH

# Key Points for Developing on the K1

- Need to remember that Android is overlaid on a Linux base
- Code development and testing (including CUDA) can be done on any PC
- Profiling on Logan
  - NVProf for Logan – can be ported to your PC

SAGIVTECH

# Key Points for Developing on the K1

- There is a strong separation between the Android system and the NDK
- A CUDA developer doesn't need to become an Android developer
- From the Android developer viewpoint this is simply a library
- An Android developer doesn't need to become a CUDA developer

SAGIVTECH

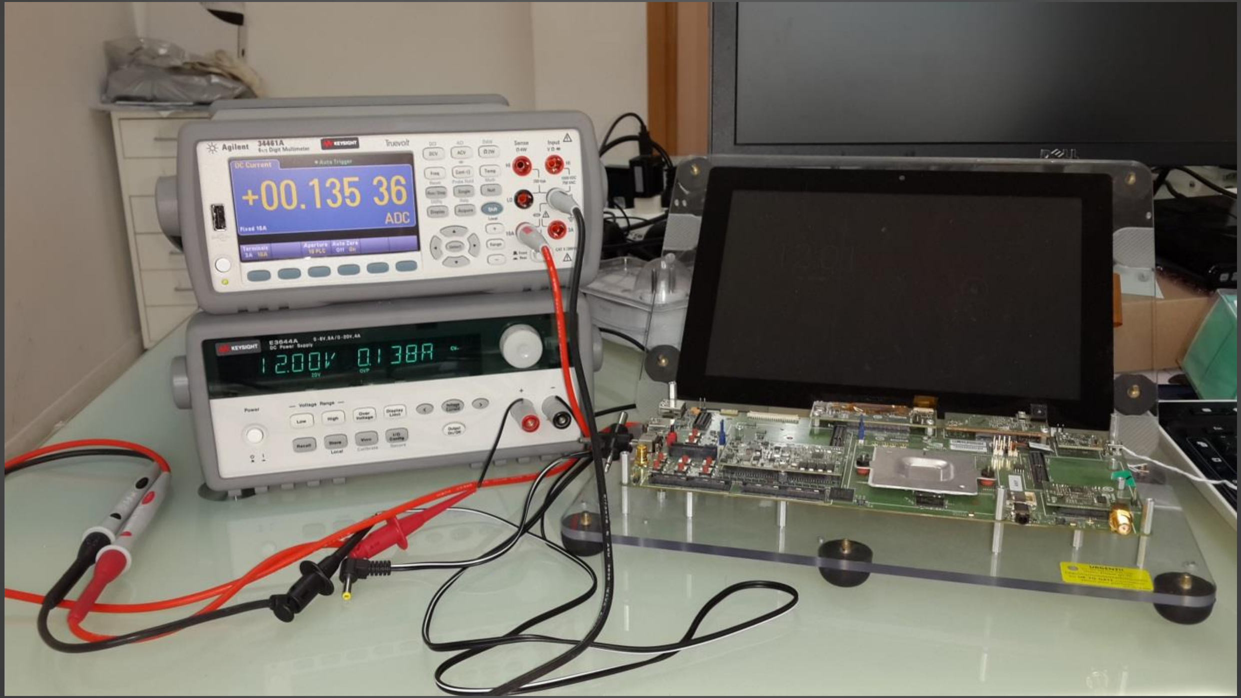
# Take Home Tips for CUDA on Tegra K1

- Development methodology is similar to discrete GPU development

- Don’t underestimate Tegra’s CPU  - the challenge is to divide work between the various components

SAGIVTECH

# Optimization of GigaFlops/WATT




SAGIVTECH

# Mobile Crowdsourcing Video Scene Reconstruction

- If you've been to a concert recently, you've probably seen how many people take videos of the event with mobile phone cameras

- Each user has only one video – taken from one angle and location and of only moderate quality

# The Idea behind SceneNet

Leverage the **power of** multiple mobile phone cameras

to create a **high-quality 3D** video experience that is

**sharable** via social networks

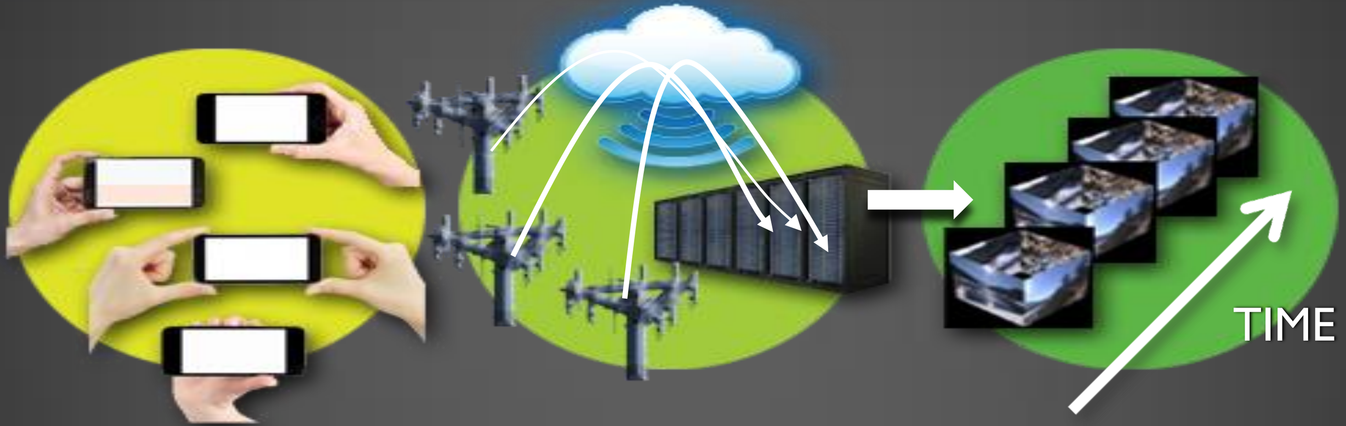
SAGIVTECH

SCENE NET

# The Combined Model: Mobile & Cloud Computing


SAGIVTECH

# Creation of the 3D Video Sequence



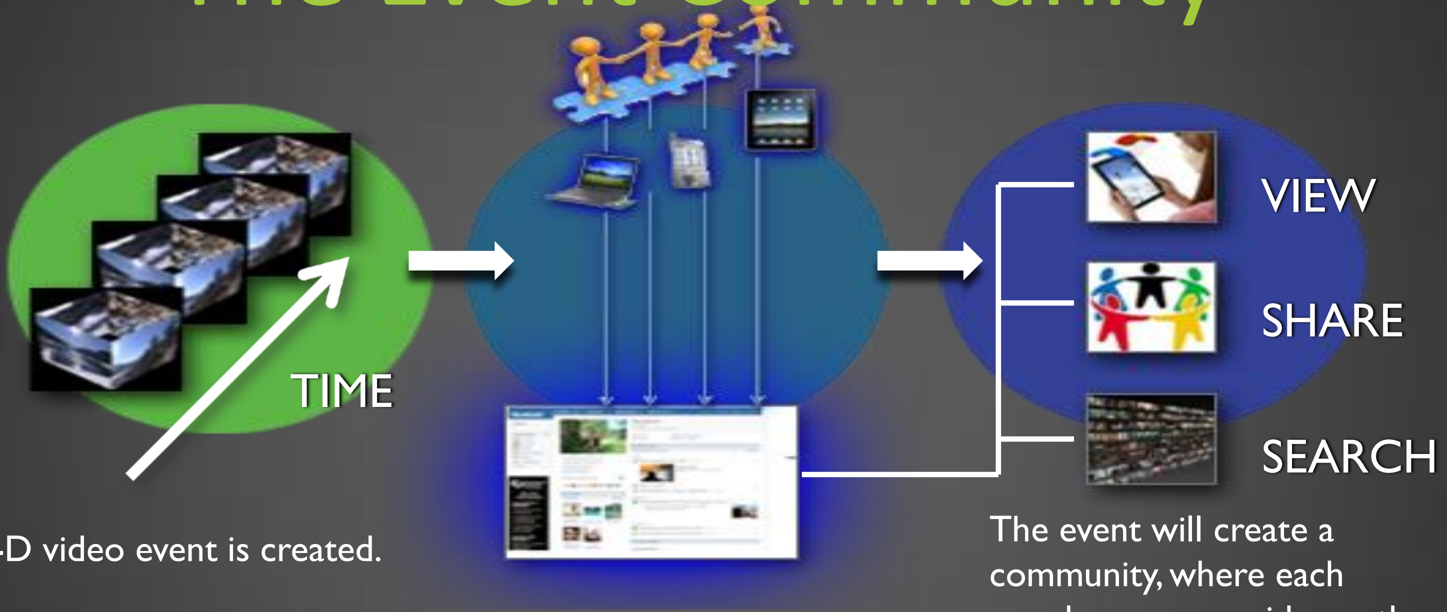

The scene is photographed by several people using their cell phone camera

The video data is transmitted via the cellular network to a High Performance Computing server.

Following time synchronization, resolution normalization and spatial registration, the several videos are merged into a 3-D video cube.

SAGIVTECH

# The Event Community

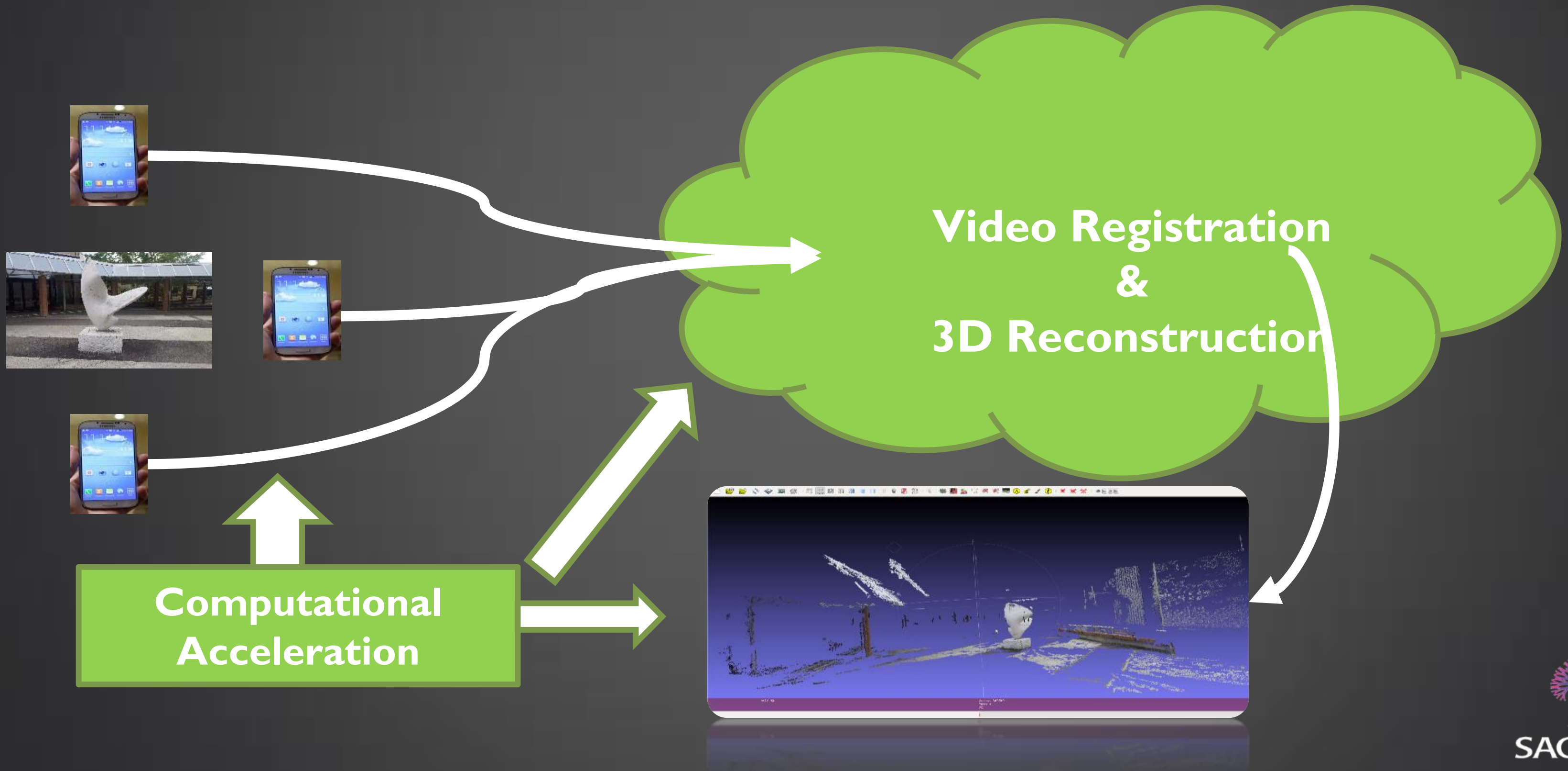



A 3-D video event is created.

The 3-D video event will be available on the internet as public or private event.

The event will create a community, where each member may provide another piece of the puzzle and view the entire information.

# GPU Computing in SceneNet



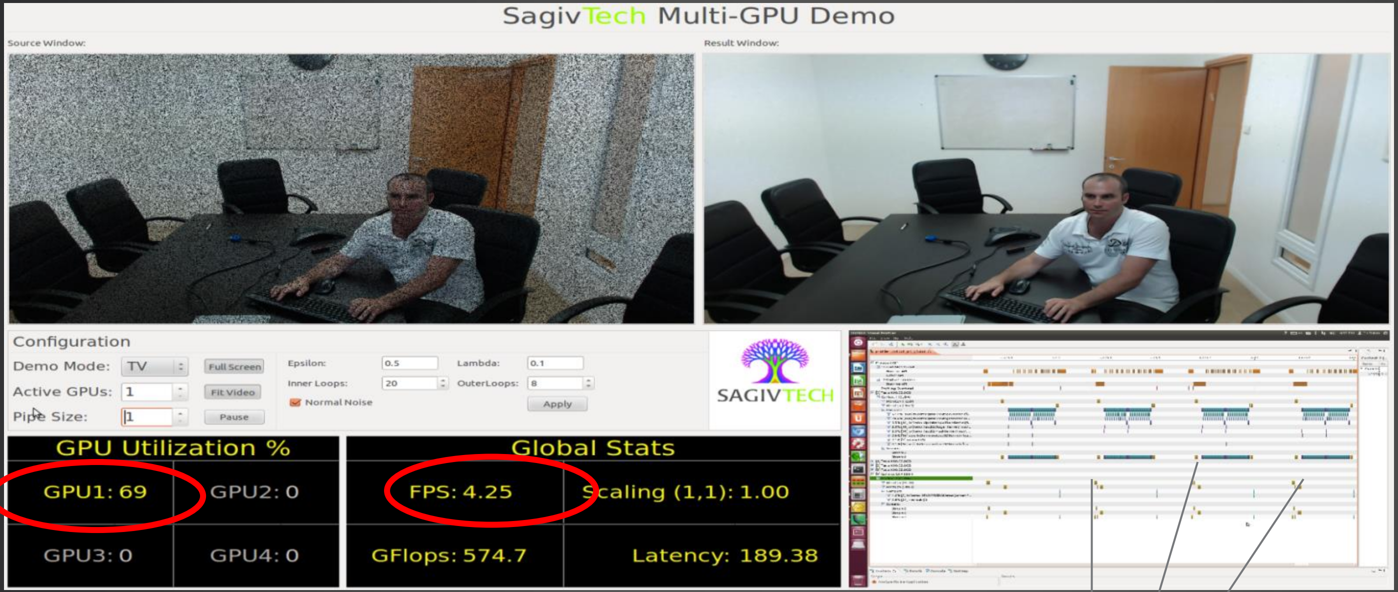

SAGIVTECH

# Mobile – Cloud Concept

- Understanding, interpretation and interaction with our surroundings via mobile device
- Demand for immense processing power for implementation of computationally-intensive algorithms in real time with low latency
- Computation tasks are divided between the device and the server
- With CUDA – it's simply easier!

SAGIVTECH

# ST MultiGPU Real World Use Case

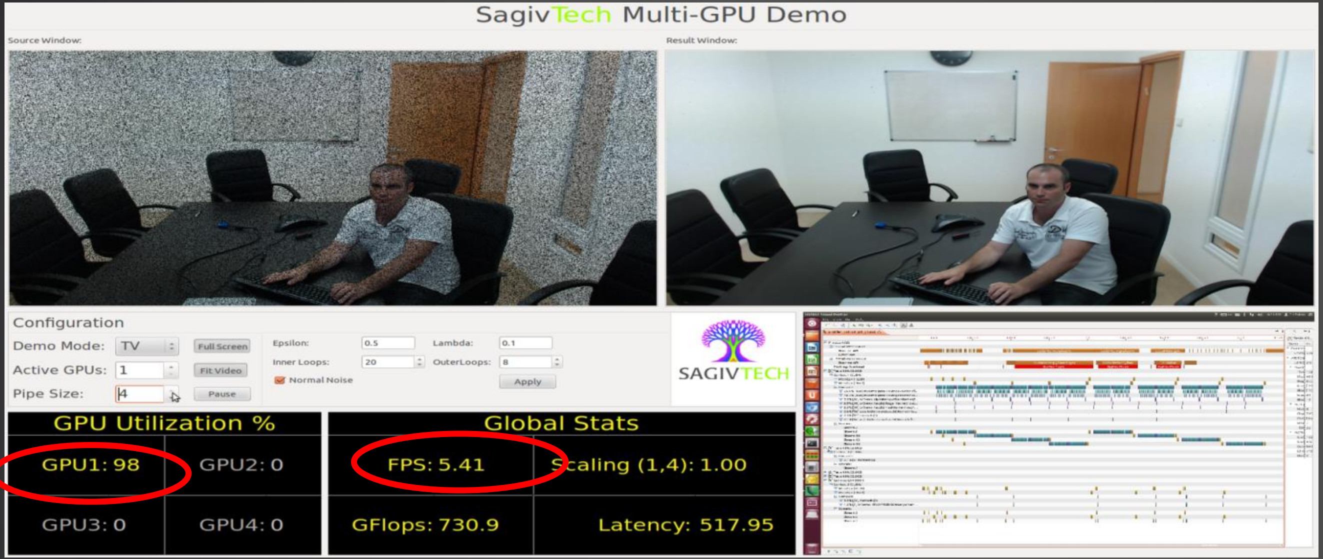



One GPU
One pipe
Utilization: ~70%

- FPS: 4.25
- Scaling: 1.00
- Note the gaps in the profiler

SAGIVTECH

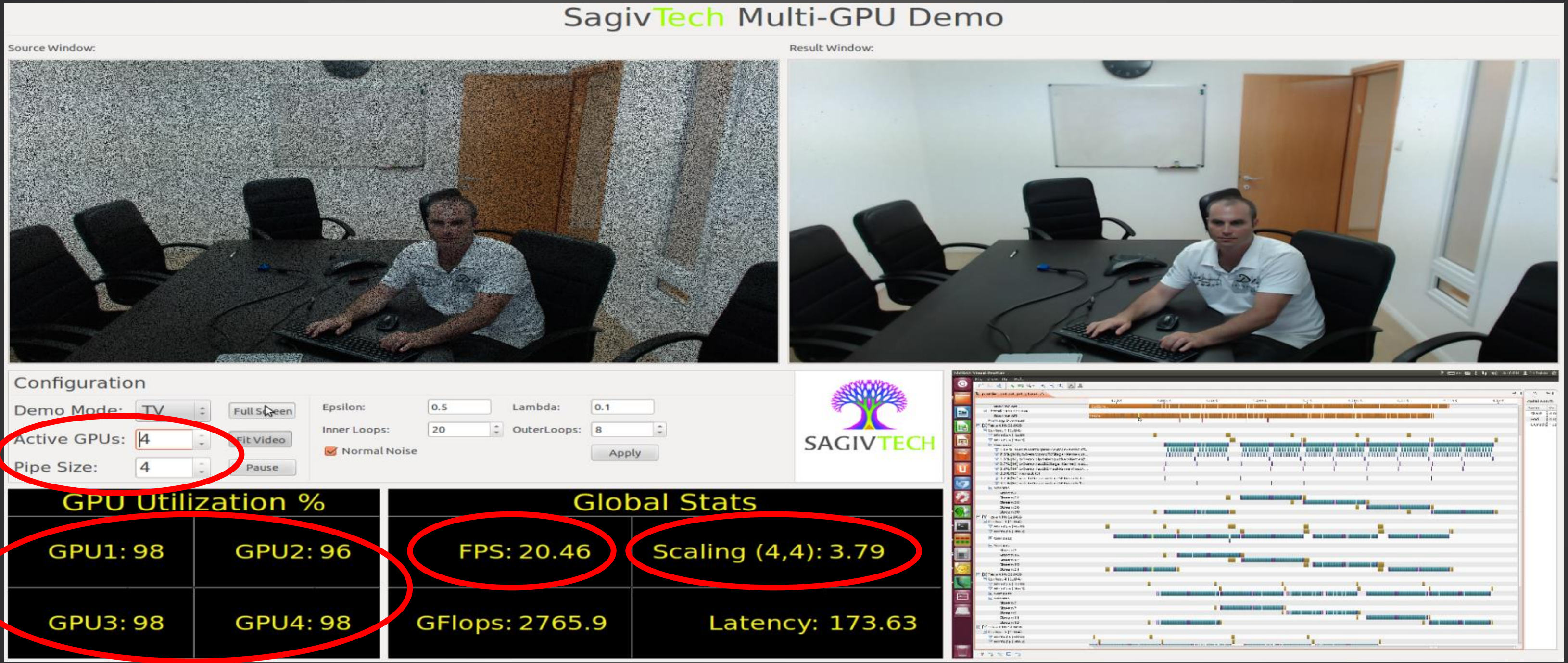
# ST MultiGPU Real World Use Case




One GPU
4 pipes
Utilization: 95%

- FPS: 5.41
- Scaling: 1.27
- Better utilization using pipes

# ST MultiGPU Real World Use Case




Four GPUs
Four pipes
Utilization: 96%+

- FPS: 20.46
- Scaling: 3.79 – Near linear Scaling!
- Note NO gaps in the profiler

# Mobile Crowdsourcing Video Scene Reconstruction

This project is partially funded by the European Union under the 7th Research Framework, programme FET-Open SME, Grant agreement no. 309169.

# Thank You

For more information please contact

Nizan Sagiv

nizan@sagivtech.com

+972 52 811 3456

SAGIVTECH