Software Self-configurability in GPU-accelerated Robot Vision

Luigi Nardi, PhD
Imperial College London
Software Performance Optimisation group

@GTC-EU Amsterdam
September 29th 2016

In collaboration with:
The three R’s of vision: Spectrum of Computer Vision Research

**Reconstruction**
Scalable Kinect Fusion (2013)

Building Rome on a cloudless day (2010)

**Recognition**
Deep learning for scalable Object class detection (2014)

**Reorganisation or Grouping**
Contour detection and segmentation (2011)

Credit: Zeeshan Zia
Simultaneous localisation and mapping (SLAM)

Build a coherent world representation and localise the camera in real-time

Sparse SLAM

Video: Dyson 360 Eye
Dense SLAM

Video: KinectFusion
[Newcombe et al. ISMAR 2011]
Simultaneous localisation and mapping (SLAM)

Build a coherent world representation and localise the camera in real-time

In this talk I will focus on two dense algorithms:

- KinectFusion [Newcombe et al. ISMAR 2011]
- ElasticFusion [Whelan et al. RSS 2015]

Applications, e.g.:
- Robotics
- Autonomous driving
- 3D printing
- Augmented reality
- Telepresence
What CV researchers say about **KinectFusion** and **ElasticFusion** performance

"Cannot run in real-time on mobile"

"You need a fat GPU to run it"
Holistic approach to SLAM performance:

**SLAMBench**

A publicly-available benchmarking framework for quantitative, comparable and validatable experimental research to investigate trade-offs in performance, accuracy and energy consumption of a SLAM system.

Error metric: absolute trajectory error (ATE) based on dataset ground truth

*Introducing SLAMBench, a performance and accuracy benchmarking methodology for SLAM (ICRA 2015)*
SLAMBench framework

SLAM benchmarks
- KinectFusion
- ElasticFusion
- LSD-SLAM
- ORB-SLAM
- Dense SLAM
- Semi-dense SLAM
- Sparse SLAM

Implementation languages
- C++
- OpenMP
- OpenCL
- CUDA
- SYCL
- PENCIL

Desktop to embedded platforms
- ARM
- Intel
- NVIDIA

Datasets
- ICL-NUIM
- TUM RGB-D

Performance evaluation
- Frame rate
- Energy
- Accuracy
**KinectFusion** algorithmic features

<table>
<thead>
<tr>
<th>Features</th>
<th>Ranges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume resolution</td>
<td>64x64x64, 128x128x128, 256x256x256, 512x512x512</td>
</tr>
<tr>
<td>$\mu$ distance</td>
<td>0 .. 0.5</td>
</tr>
<tr>
<td>Pyramid level iterations (3 levels)</td>
<td>0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11</td>
</tr>
<tr>
<td><strong>Image resolution (image ratio)</strong></td>
<td>1, 2, 4, 8</td>
</tr>
<tr>
<td>Tracking rate</td>
<td>1, 2, 3, 4, 5</td>
</tr>
<tr>
<td>ICP threshold</td>
<td>$10^{-6} .. 10^2$</td>
</tr>
<tr>
<td>Integration rate</td>
<td>1 .. 30</td>
</tr>
</tbody>
</table>

Different algorithmic features for **ElasticFusion**
Motivation

• KinectFusion runtime response surface: non-linear, multi-modal and non-smooth
• Optimal **algorithm configurability** enables better performance and better accuracy of the computation

*Integrating Algorithmic Parameters into Benchmarking and Design Space Exploration in 3D Scene Understanding* (PACT 2016)
Exploration goal illustration

- Targeted prediction area
- Pareto front
- Samples
- Error
- Runtime threshold
- Accuracy threshold
Algo Design-Space Exploration (DSE): Active Learning Methodology

Samples

Algorithmic configuration parameters

Machine learning

Active learning

Predictive model

Time Error

Run
Machine learning methods used

**Decision Tree**

**Random Forest**
## Results

### ElasticFusion DSE error/runtime III

<table>
<thead>
<tr>
<th>Machine</th>
<th>Type</th>
<th>CPU</th>
<th>CPU name</th>
<th>CPU cores</th>
<th>GPU</th>
<th>GPU name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVIDIA/Intel</td>
<td>Desktop</td>
<td>Intel</td>
<td>Ivy Bridge</td>
<td>E5-1620</td>
<td>8</td>
<td>NVIDIA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GTX 780 Ti</td>
</tr>
</tbody>
</table>

![Graph showing mean time per frame (sec) vs. max ATE (m)]

- **Default configuration**
- **Active learning**
- **Random sampling**

Accuracy limit = 0.05m
Results ElasticFusion DSE error/runtime III

<table>
<thead>
<tr>
<th>Machine</th>
<th>Type</th>
<th>CPU</th>
<th>CPU name</th>
<th>CPU cores</th>
<th>GPU</th>
<th>GPU name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVIDIA/Intel</td>
<td>Desktop</td>
<td>Intel</td>
<td>Ivy Bridge</td>
<td>E5-1620</td>
<td>8</td>
<td>NVIDIA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GTX 780 Ti</td>
</tr>
</tbody>
</table>

Max ATE (m) - Accuracy limit = 0.05m

- Default configuration
- Active learning
- Random sampling

- 45 FPS, 0.056 ATE
- 65 FPS, 0.0332 ATE
- 58 FPS, 0.026 ATE
Results **KinectFusion** DSE error/runtime I

<table>
<thead>
<tr>
<th>Machine</th>
<th>Type</th>
<th>CPU</th>
<th>CPU name</th>
<th>CPU cores</th>
<th>GPU</th>
<th>GPU name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardkernel ODROID-XU3</td>
<td>Embedded</td>
<td>ARM A15 + A7</td>
<td>Exynos 5422</td>
<td>4 + 4</td>
<td>ARM</td>
<td>Mali-T628</td>
</tr>
</tbody>
</table>

Accuracy limit = 0.05m

- Default configuration
- Active learning
- Random sampling
Results **KinectFusion** DSE error/runtime I

<table>
<thead>
<tr>
<th>Machine</th>
<th>Type</th>
<th>CPU</th>
<th>CPU name</th>
<th>CPU cores</th>
<th>GPU</th>
<th>GPU name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardkernel ODROID-XU3</td>
<td>Embedded</td>
<td>ARM A15 + A7</td>
<td>Exynos 5422</td>
<td>4 + 4</td>
<td>ARM</td>
<td>Mali-T628</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean time per frame (sec)</th>
<th>Max ATE (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.035</td>
<td>0.044 ATE</td>
</tr>
<tr>
<td>0.040</td>
<td>0.046 ATE</td>
</tr>
<tr>
<td>0.045</td>
<td>0.041 ATE</td>
</tr>
<tr>
<td>0.050</td>
<td>0.044 ATE</td>
</tr>
<tr>
<td>0.055</td>
<td>0.046 ATE</td>
</tr>
</tbody>
</table>

- Default configuration
- Active learning
- Random sampling

Accuracy limit = 0.05 m

6 FPS, 0.044 ATE

38 FPS, 0.046 ATE

3 FPS, 0.041 ATE
Results **KinectFusion** DSE error/runtime II

<table>
<thead>
<tr>
<th>Machine</th>
<th>Type</th>
<th>CPU</th>
<th>CPU name</th>
<th>CPU cores</th>
<th>GPU</th>
<th>GPU name</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASUS T200TA</strong></td>
<td>Detachable laptop</td>
<td>Intel Silvermont</td>
<td>Atom Z3795</td>
<td>4</td>
<td>Intel</td>
<td>HD Graphics</td>
</tr>
</tbody>
</table>

- **Default configuration**
- **Active learning**
- **Random sampling**

Accuracy limit = 0.05m

Max ATE (m) vs. Mean time per frame (sec)
Conclusion - take away messages

1. Building tools to explore the performance landscape for SLAM solutions
2. Pareto maps how configurations should be adapted when objectives change - static and dynamic
3. Performance and accuracy improvement over default
4. Generalisation to other applications
Live demos GTC-EU 2016

Imperial College London

Booth D3

Passenger Terminal Hall 1/Main deck

September 28th and 29th
SLAMBench install
Booth D3 for help

Publicly released 13/11/2014
(1400+ downloads)
References I

References II


Copyrights

- Author: Dyson Ltd. Dyson 360 Eye. [Video]. Retrieved from https://www.youtube.com/watch?v=OadhuICDAjk