3D Graphics System Challenges for Simulation: Lessons from AI Habitat

Manolis Savva

HPG 2020
2020-07-16
Preface: “thoughts from a graphics expat”
“Simulation”?
Terminology: Embodied AI

"The embodiment hypothesis is the idea that intelligence emerges in the interaction of an agent with an environment and as a result of sensorimotor activity."

The Development of Embodied Cognition: Six Lessons from Babies [Smith & Gasser 2005]
Embodied Agents

Physically embodied agents taking actions in the world

= Human-like AI
  • Active perception
  • Long-term planning
  • Learning by interaction

Image credits: DRC-Hubo robot [DARPA Robotics Challenge], [Adrian Murray / Trevillion Images]
Simulation for embodied AI

Physically embodied agents taking actions in the world

Virtual embodied agents taking actions in a virtual world
Internet AI → Embodied AI

Image Credit: Image-Net
Slide credit: Abhishek Das
Image Credit: Lockheed Martin; DARPA Robotics Challenge
From internet image datasets to 3D simulators

Dataset $\rightarrow$ Simulator $\rightarrow$ Task $\rightarrow$ Benchmark
Year 2017: exciting times!
3D simulators galore!

HoME Platform [Brodeur et al. 2017]

House3D [Wu et al. 2017]

MINOS [Savva et al. 2017]

AI2-THOR [Kolve et al. 2017]

Matterport3D Simulator [Anderson et al. 2018]

Gibson Environment [Zamir et al. 2018]

InteriorNet / ViSim [Li et al. 2018]
## 3D simulators galore!

<table>
<thead>
<tr>
<th>Environment</th>
<th>3D</th>
<th>Large-Scale</th>
<th>Customizable</th>
<th>Physics</th>
<th>Photorealistic</th>
<th>Actionable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atari</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OpenAI Universe</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malmo</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeepMind Lab</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VizDoom</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matterport3D</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>MINOS (Matterport3D)</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>House3D</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MINOS (SUNCG)</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HoME</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AI2-THOR</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Table from AI2-THOR [Kolve et al. 2017]
Impact: research tasks and communities

Visual navigation

Instruction following

Robotic manipulation

[Anderson et al. 2018]

[Gupta et al. 2017]

[James et al. 2019]
Common: black-boxed 3D game engine binary

AI2-THOR [Kolve et al. 2017]
architecture example sketch

10 – 60 FPS
However: not for human eyeballs!

**Human:** 1080p @ 60Hz

**RL:** 84x84 @ 1000+ Hz
Can we do better?
Habitat: A Platform for Embodied AI Research

aihabitat.org

Manolis Savva*  Abhishek Kadian*  Oleksandr Maksymets*  Yili Zhao  Erik Wijmans  Bhavana Jain

Julian Straub  Jia Liu  Vladlen Koltun  Jitendra Malik  Devi Parikh  Dhruv Batra

[Logos of various institutions: Facebook Artificial Intelligence Research, Facebook Reality Labs, Georgia Tech, SFU, Intel, Berkeley University of California]
Habitat: standardizing the Embodied AI “software stack”

**Simulators**
- **House3D** (Wu et al., 2017)
- **AI2-THOR** (Kolve et al., 2017)
- **MINOS** (Savva et al., 2017)
- **Gibson** (Zamir et al., 2018)
- **CHALET** (Yan et al., 2018)

**Tasks**
- EmbodiedQA (Das et al., 2018)
- Language grounding (Hill et al., 2017)
- Interactive QA (Gordon et al., 2018)
- Instruction following (Anderson et al., 2018)
- Visual Navigation (Zhu et al., 2017, Gupta et al., 2017)

**Datasets**
- Replica (Straub et al., 2019)
- Matterport3D (Chang et al., 2017)
- 2D-3D-S (Armeni et al., 2017)
Attention to speed
Did speed matter?
Learned vs classical navigation agents

To Learn or Not to Learn: Analyzing the Role of Learning for Navigation in Virtual Environments

Noriyuki Kojima
University of Michigan
2260 Hayward St, Ann Arbor, MI 48109
kojinano@umich.edu

Jia Deng
Princeton University
35 Olden St 423, Princeton, NJ 08540
jiadeng@cs.princeton.edu

Abstract

In this paper we focus on the task of geometric navigation, i.e., navigation when ground-truth 3D information is available. Specifically, we explore the dichotomy between "learning" and "coding" for this task. We construct a hand-coded navigating agent, and demonstrate that it outperforms state-of-the-art learning based agents on two popular benchmarks, MINOS [27] and Stanford large-scale 3D Indoor Spaces (53DIS) [2]. We also observe that as the environment becomes more challenging, the performance gap between learning-based and hand-coded agent increases.

We construct a hand-coded agent for the task of geometric navigation and compare its performance with state-of-the-art learning based methods on two challenging benchmarks: 53DIS [2] and MINOS [27]. On MINOS, the UNREAL agent [27] (which is based on deep reinforcement learning) and on 53DIS, the CMP agent [25] (which uses imitation learning to jointly train a mapper and plan-
Example navigation episodes

Blind Agent

Depth Agent
Back to today: simulators galore part 2!

- **RLBench** [James et al. 2019]
- **IKEA Furniture Assembly** [Lee et al. 2019]
- **SAPIEN** [Xiang et al. 2020]
- **iGibson** [Xia et al. 2020]
Emerging trends
Emerging trends: interaction

iGibson [Xia et al. 2020]

RLBench [James et al. 2019]
Emerging trends: scale (& more speed)

DD-PPO: Learning Near-Perfect PointGoal Navigators from 2.5 Billion Frames [Wijmans et al. 2020]

Sample Factory: Egocentric 3D Control from Pixels at 100000 FPS with Asynchronous Reinforcement Learning [Petrenko et al. 2020]
Emerging trends: multimodality

Audio-Visual Embodied Navigation [Chen et al. 2020]
Emerging trends: Sim2Real

RoboTHOR [Deitke et al. 2020]

Sim2Real Coefficient [Kadian et al. 2020]
Graphics system challenges
Challenge: “fast physics”

iGibson [Xia et al. 2020]  RLBench [James et al. 2019]
Challenge: “GPU cohabitation”

Rendering

Physics

Learning

OpenGL

Vulkan

NVIDIA

PyTorch

TensorFlow
Challenge: “not for eyeballs”

**Human:** 1080p @ 60Hz

**RL:** 84x84 @ 1000+ Hz
Challenge: “asset soup”

**AI2-THOR**
120 virtual rooms

**Replica**
18 near-photorealistic rooms

**ShapeNet**
65K virtual objects

**Matterport3D**
90 multi-floor house reconstructions

AI2-THOR [Kolve et al. 2017], Replica [Straub et al. 2019], ShapeNet [Chang et al. 2016], Matterport3D [Chang et al. 2017]
Summary

Trends
• Interaction
• Scale & more speed
• Multimodality
• Sim2Real

Challenges
• “Fast physics”
• “GPU cohabitation”
• “Not for eyeballs”
• “Asset soup”
Takeaway messages

• Growing interest in embodied AI

• Simulation for embodied AI: new frontiers for GFX-ML systems

• Opportunities for broad impact!
Visual Computing @ Simon Fraser University

We’re hiring at all levels! MSc, PhD, postdocs, researchers, faculty 😊

msavva@sfu.ca
Thank you!