Keynote at ICONS 2021

Human Skill Transfer for AI Accelerator Programming and Its Application to Humanoid Two-Handed Robots

Prof. Chung-Ta King
king@cs.nthu.edu.tw
Department of Computer Science
National Tsing Hua University, Taiwan
About Me

• 1988: Ph.D., Michigan State University, USA
• 1988-1990: Assistant Professor, NJIT, USA
• 1990-1997: Associate Professor, NTHU, Taiwan
• 1997-: Professor, NTHU
• 2009-2012: Chair, Dept of CS, NTHU
• 2014-2015: Associate Dean, College of EECS, NTHU
• 2019-: Vice President and Chief of Staff, NTHU
• Research interests: embedded & mobile computing, computer architecture, parallel & distributed systems
• Conference organizer: Cluster 2016, ICPADS 2014, CloudCom 2012, ESWEEK 2011, ...
Acknowledgements

• This talk is based the outcomes obtained so far from a 4-year project (2019-2022) funded by the Minister of Science and Technology, Taiwan, under the Moon-Shoot Project

• Team members:
  – Professors Cheng-Wen Wu, Jen-Yuan Chang, Jing-Jia Liou, Chih-Tsun Huang, and Hung-Kuo Chu of National Tsing Hua University, Taiwan
  – Over 20 research staffs and graduate students
Growing Importance of Service Robots

Challenge: Complexity in robot controls due to very high *degree-of-freedom* (DOF) and complex working environments ➔ application developments

Medical care, logistics, farming, household, companion, education, entertainment, ...

Growing market size to > $20 Billion in 2024

Humanoid, 5-fingered hands desirable for adapting to varying tasks and dynamic environments

https://www.pinterest.com/pin/127719339416902130/
Fast Development of Robot Applications

Common practices:
- Programming by programmers
- Moving arms to set actions
- Training and imitation

To handle dynamics:
- Sensing: vision, touching
- Planning and decision
- Control
Imitation for Fast Application Development

Human demonstrates:
- What are sensed (vision, touch, ...)?
- What actions are performed in response?
→ Without expertise in robotics

Robot imitates:
- Not just clone trajectories and actions (behavior cloning)
- Must be able to adapt and generalize

Reinforcement/Imitation learning
AI accelerator
Problems with Imitation

- Scattered tools relying on manual interfacing
- Training with physical robots too slow and unsafe
- Process from NN models to AI accelerators too long
- Design space exploitation difficult and inefficient

Need a systematic flow for fast development of robot applications -- from human demonstration onto accelerator chips
Human Skill Transfer for AI Accelerators

Two design flows:

• From human skills to AI model for robot control
  – Collect data from human demonstration to train robot control models in a virtual environment

• From AI model to AI accelerator chips
  – Compile and optimize the AI control model to run on AI accelerators

A humanoid, dual-hand robot:

• Controlled by AI accelerators that run neural networks
Project Overview

- Demonstration data collection
- Virtual training environment
- Virtual Design Platform
  - Ctrl
  - Mem
  - PEs
- Robot control accelerator
- Humanoid dual-hand robot
  - Obj detect SNN chip
  - RGB-D camera
- High DoF arm
- Soft hand
- Virtual-physical AI accelerator development platform
- Imitation-based robot training platform
- Demonstration data collection

National Tsing Hua University
Demonstration of Trained Robot Operations

Operations by physical robot (3X speed)

Operations in simulator
Demonstration of Trained Robot Operations

Operations by physical robot (3X speed)

Operations in simulator
Humanoid Dual-Hand Robot

(Prof. Jen-Yuan Chang)

7-DoF arms
- Redundant DoF kinematics
- 7-DoF obstacle avoidance

Vision system
- RGB-D camera
- 2-DoF head rotation

Wire-drive 5-fingered hand
- Underactuated wire-driven 5-fingered gripper
- Touch sensor feedback

Soft hand + IMU
- Flexible 5-fingered gripper
- Pneumatic actuation
- IMU for angle feedback

Working Space

Obstacle Avoidance
Self-Adaptive Finger Joints

(Prof. Jen-Yuan Chang)

Underactuated design:
- 1 actuator to drive two joints
- Better grasps of unknown objects
- Self-adaptive to errors in grasping

<table>
<thead>
<tr>
<th></th>
<th>Shadow Robot – Hand Lite</th>
<th>NTHU Hand</th>
</tr>
</thead>
<tbody>
<tr>
<td>DoF</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Weight</td>
<td>2.4 kg</td>
<td>1 kg</td>
</tr>
<tr>
<td>Loading</td>
<td>4 kg</td>
<td>1 kg → 3 kg</td>
</tr>
<tr>
<td>Driving</td>
<td>Wired (motor fixed)</td>
<td>Wired (motor separated)</td>
</tr>
</tbody>
</table>

- Underactuated design: 1 actuator to drive two joints
- Better grasps of unknown objects
- Self-adaptive to errors in grasping
Magnetic-based Touch Sensor

(Prof. Jen-Yuan Chang)
Softhand + IMU

(Complementary Unscented Kalman Filter, CUKF)

- Gyro update
  - IMU Info.
- Acc Mag update
- Display

(Prof. Jen-Yuan Chang)
Real-time Obstacle Avoidance

Spring Vector Repulsive Force Algorithm

- Obstacle Velocity
- Repulsive Force
- Obstacle
- Robot End Effector

Object and skeleton detection

(Prof. Jen-Yuan Chang)
Robot Vision System

Rotate camera to increase field-of-view

Depth of transparent objects

- Generate Input image (130ms)
- Define problem & equation & build matrix (170ms)
- Use AMGCL library to accelerate (226ms → 8ms)
- Write image & read image (100ms)

(Robot Vision System)

ClearGrasp [1]

(Prof. Hung-Kuo Chu)
Object Detection SNN AI Chip

Application-specific NN-based AI accelerator design:

- Choose network parameters to optimize hardware cost and power consumption in terms of accuracy, network throughput, ...

Spiking Neural Network (SNN)
Object Detection SNN AI Chip

(Prof. Cheng-Wen Wu)

- Based on spiking neural network
- Application-specific (label detection)
- Optimized parameters

<table>
<thead>
<tr>
<th>Design</th>
<th>Ours</th>
<th>BinarEye CICC’18</th>
<th>IBM TrueNorth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>90nm</td>
<td>28nm</td>
<td>28nm</td>
</tr>
<tr>
<td>Area [mm²]</td>
<td>2.07</td>
<td>1.4</td>
<td>430</td>
</tr>
<tr>
<td>Algorithm</td>
<td>SNN</td>
<td>DNN</td>
<td>SNN</td>
</tr>
<tr>
<td>Voltage [V]</td>
<td>0.6 – 1</td>
<td>0.66 – 0.9</td>
<td>1</td>
</tr>
<tr>
<td>Frequency [MHz]</td>
<td>10 – 100</td>
<td>1.5 – 48</td>
<td>-</td>
</tr>
<tr>
<td>Efficiency* [TOPS/W]</td>
<td>331 – 45</td>
<td>230 – 145</td>
<td>-</td>
</tr>
<tr>
<td>MNIST Energy* [uJ/inference]</td>
<td>0.18 @ 98.01%</td>
<td>0.2 @ 97.4%</td>
<td>6.5 @ 97.5%</td>
</tr>
</tbody>
</table>

*values are scaled to 28nm technology node.

1.4X efficiency, 10% less power
Object Detection SNN AI Chip

(Prof. Cheng-Wen Wu)

Control API

Chip demo platform

- 300 detections/sec
- 0.6 V operating voltage

FPS: 29.00

Chip demo platform

USB
Imitation-based Training Platform

Demo data collection

- Demonstrator looked at the virtual scenes to perform actions

Training in simulator

Webots

Imitation Learning

Control

NN model

Record

TVM compiler

Robot control code

(Prof. Chung-Ta King)
Electronics-Mechatronics Co-Simulation

(Prof. Jing-Jia Liou)

- AI models
- TVM-based AI Compiler
- Control code

Virtual Design Platform

- Cntrl
- Mem
- PEs

Electronics-Mechatronics Co-Simulation

- Sensor
- Control
- HW delays

Robot Simulator

- SystemC and Webots data exchange interface
- Control hardware delays are feedback to robot arm
- Verify accuracies of AI models on robot arm operations
Architecture of Virtual Design Platform (Prof. Jing-Jia Liou)

- Hardware components:
  - MCU, DMA
  - AI accelerator modules: Conv, Pool, Relu, etc.
  - Shared memory of AI accelerator modules
- RISC-V MCU for controlling dataflows of various AI accelerator modules
- All in SystemC
**Electronics-Mechatronics Co-Simulation** (Prof. Jing-Jia Liou)

**AI model for robot arm**
- Grasping
- 4 FC layers (64, 32, 16, 7)

**TVM optimization**
- Pipelined execution
- Operator fusion
- 14%X speedup

**Data exchange**
- Delays of accelerator hardware to robot simulator
Electronics-Mechatronics Co-Optimization

(Prof. Chih-Tsun Huang)
Layer Fusion for Data Reuse in Acc. Memory

• For Depthwise Separable CNNs

• Evaluation for MobileNet-v2
  • 67% speed improvement
  • 65% power consumption improvement
Conclusions

- An AI chip development environment based on human skill transfer
  - Automate the process and lower the complexity in developing service robot applications
  - Humanoid robot hand with 7-DOF arms, CMOS spiking neural network ASIC, design flows from human skills to AI model and to AI accelerator for robot control
  - Allow more complex robot applications to be more easily developed at a lower cost

- On-going works
  - Robot control with touch sensor feedbacks, memristor-based SNN, long-horizon task planning, hierarchical control models, ...