Early Implementation Experience with Wearable Cognitive Assistance Applications

Zhuo Chen, Lu Jiang, Wenlu Hu, Kiryong Ha, Brandon Amos,
Padmanabhan Pillai, Alex Hauptmann, and Mahadev Satyanarayanan

Carnegie Mellon University

intel

Wearable Cognitive Assistance

Generalize metaphor from GPS

Input:
your destination

Guidance:

- step by step directions
- know your location
- “Recalculating…”

Wearable Cognitive Assistance

Generalize metaphor from GPS

Input:
some target *task*

Guidance:

- step by step *instructions*
- know your *progress*
- *corrective feedback*

Cognitive assistance is so broad a concept

Focus on narrow, well-defined *task assistance* for now

Real World Use Cases

Industrial Troubleshooting

Medical Training

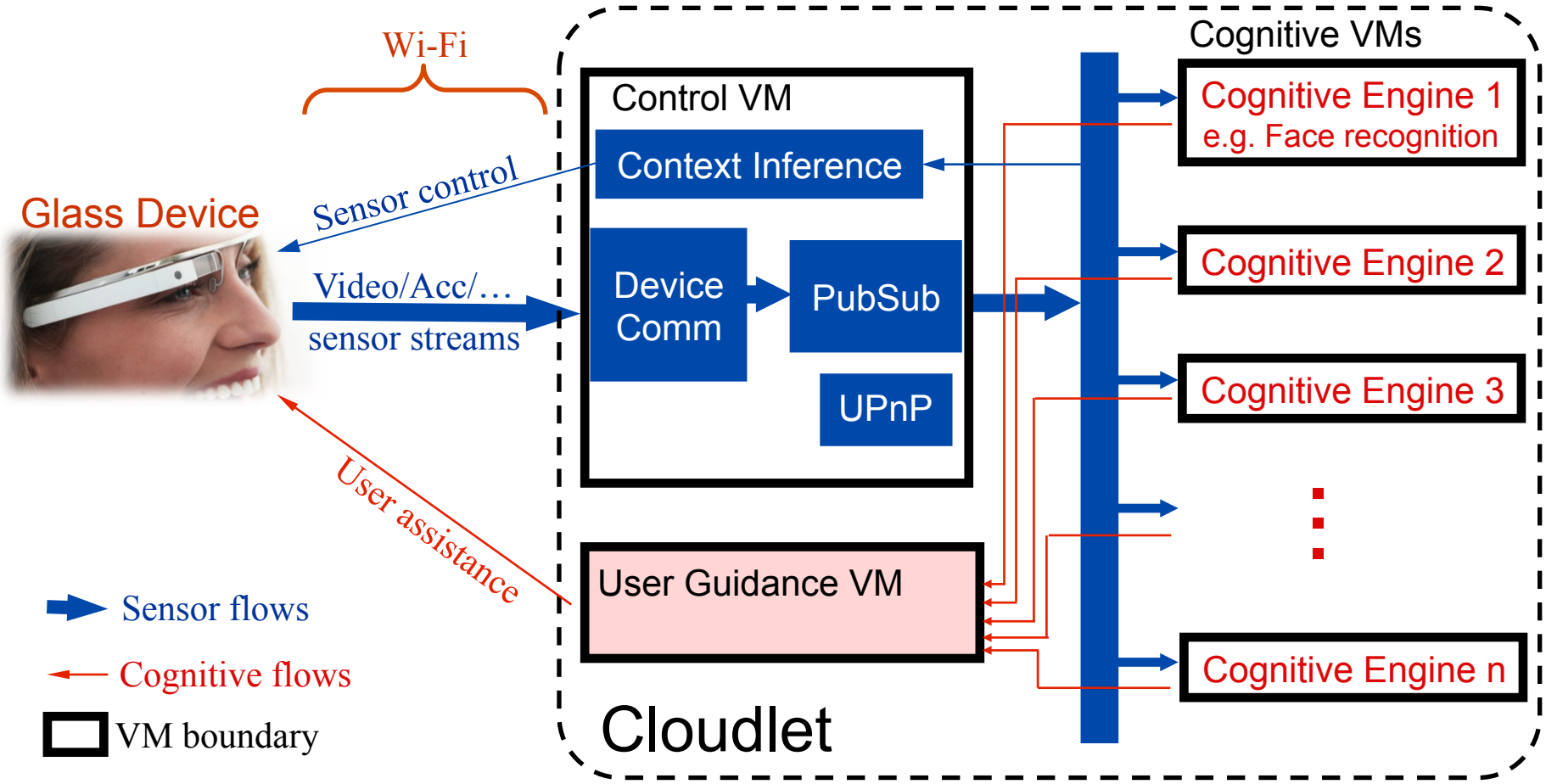
Cooking

Furniture Assembly

This Paper

- Common platform - Gabriel
- Current implementations
  - Lego Assistant
  - Drawing Assistant
  - Ping-pong Assistant
  - Assistance with Crowed-sourced Tutorials
- Lessons and Future Work

Review of Gabriel



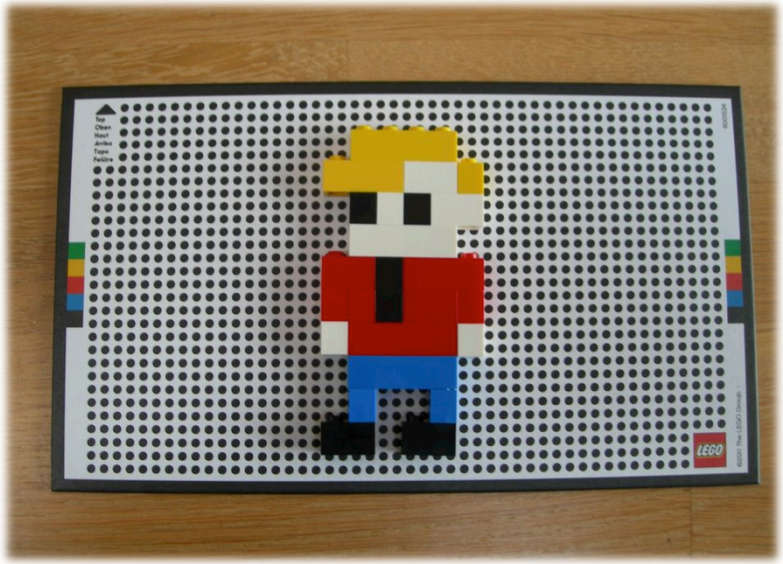
Key Features of Gabriel

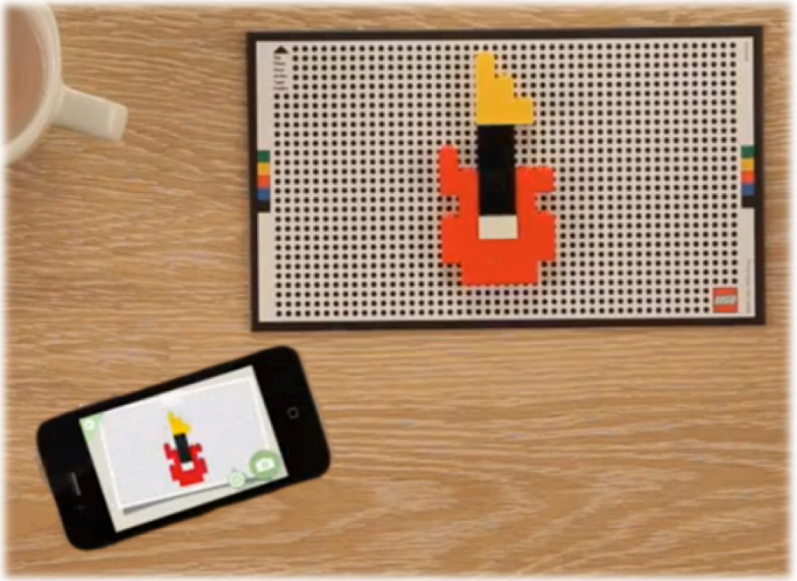
- Offload video stream to cloudlet
  - Guarantees low latency with app-level flow control
- Encapsulate each application into a VM
  - Use Pub-Sub to distribute streams

Goal: provide common functionalities to simplify development of each application

Example 1: Lego Assistant

Assembly 2D Lego with *Life of George*

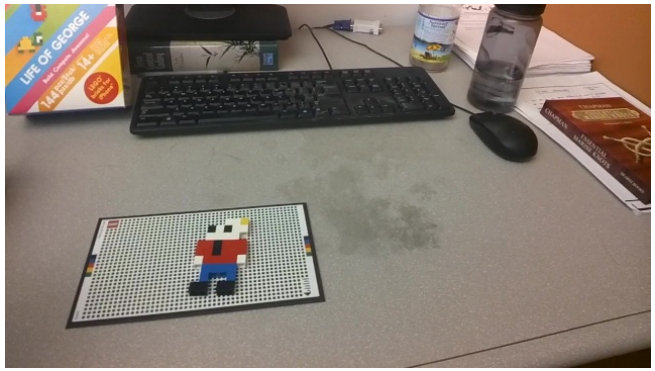




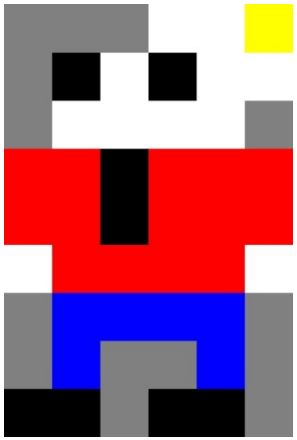
Two-phase Processing

Applies to all applications we have built

Raw stream



Symbolic Representation

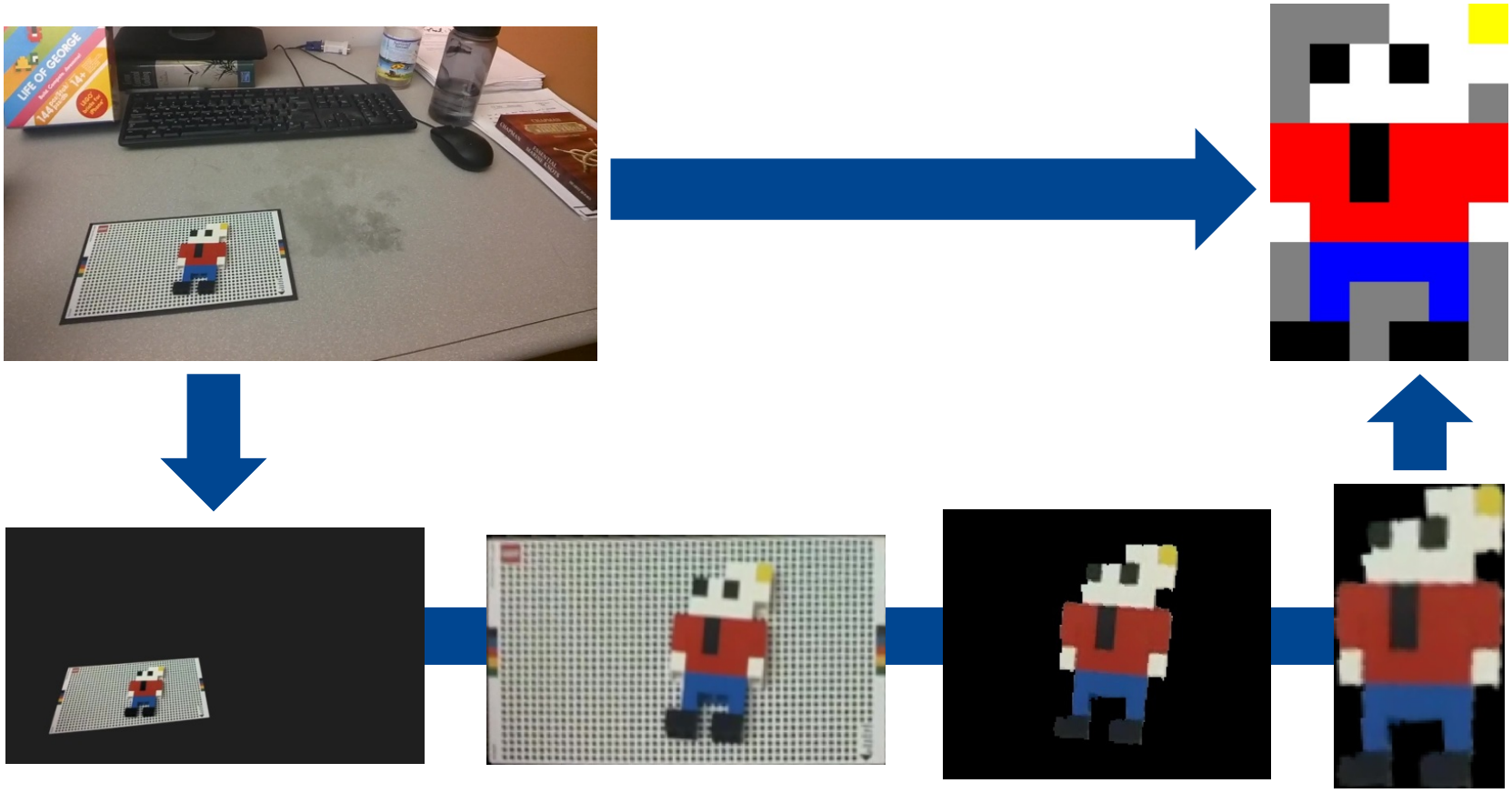


Visual + Verbal Guidance

- **“Digitize”**
- **Tolerant of different lighting, background, occlusion**

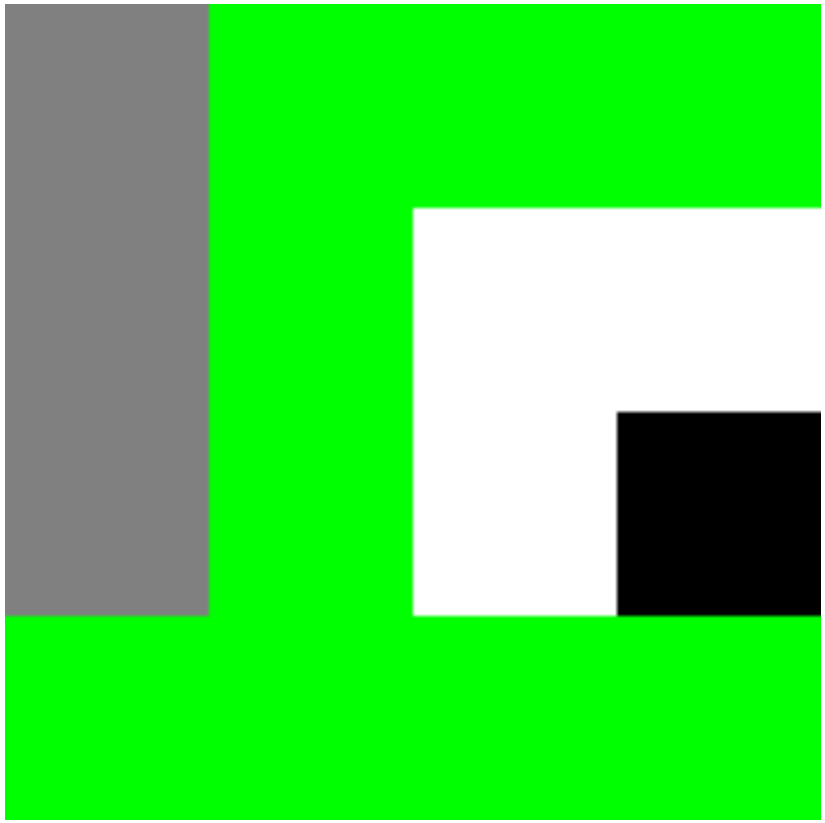
Match current state with all known states in DB to get guidance

Lego: Symbolic Representation Extractor

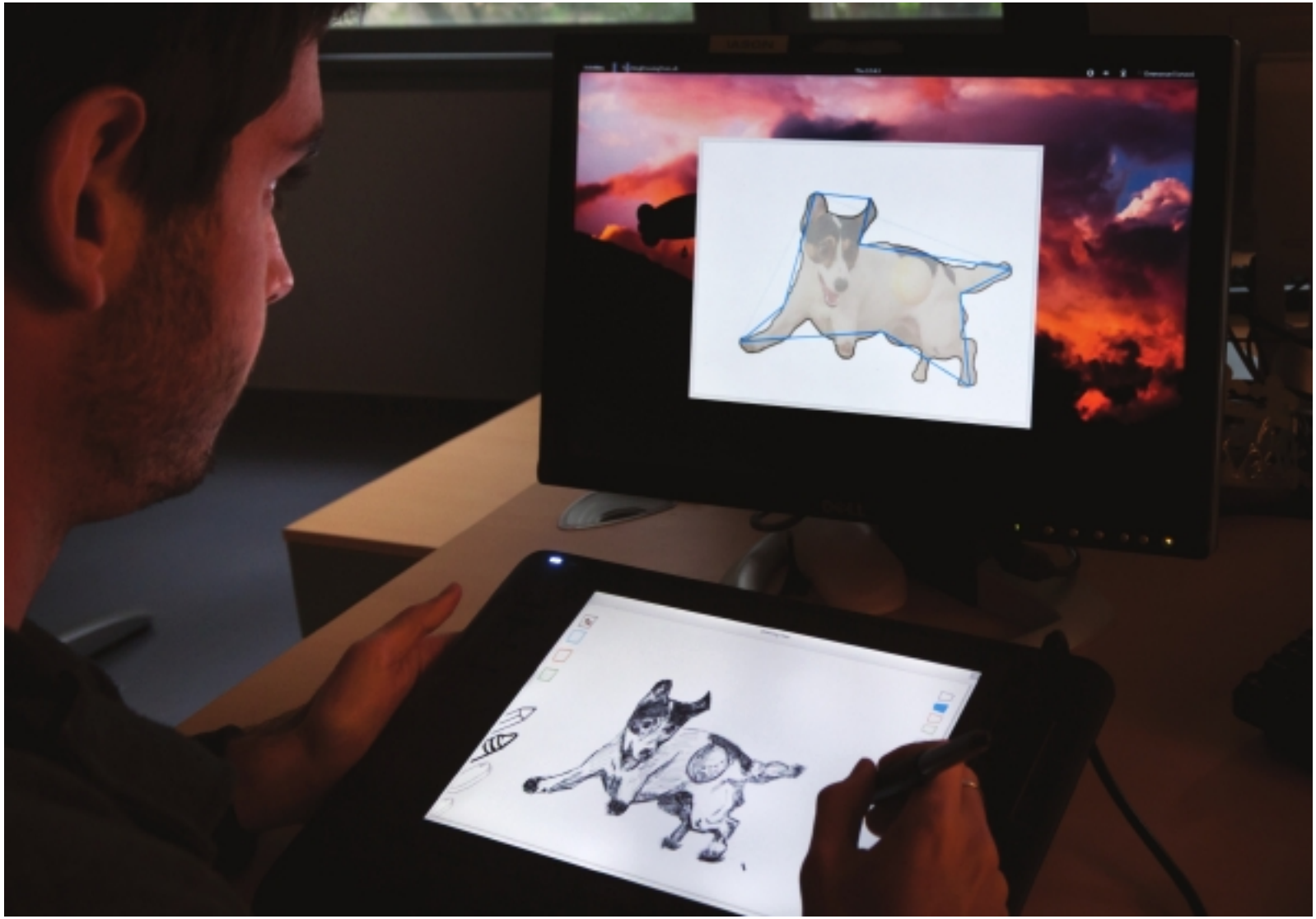


Four months of effort to make it *robust*
Spent a great amount of time on tuning parameters and testing

Lego Assistant Demo

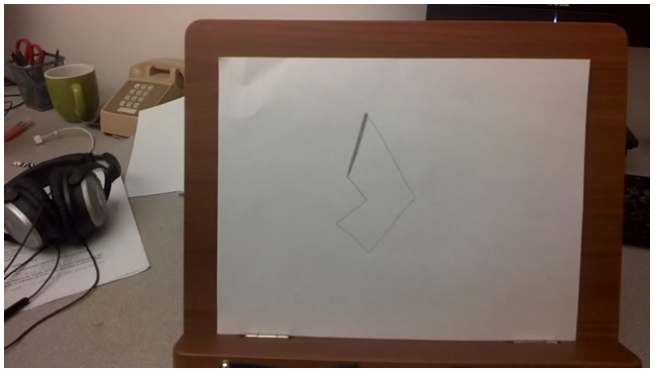


Example 2: Drawing Assistant

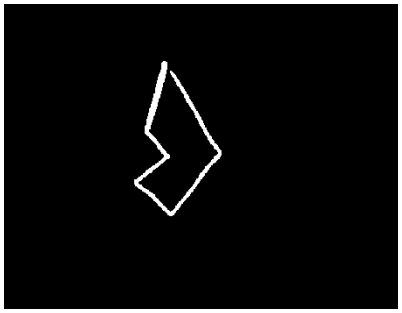


Drawing Assistant Workflow

Raw stream



Symbolic Representation (binary image)



Visual Feedback

- **Find paper**
- **Locate sketches**
- **Remove noise**

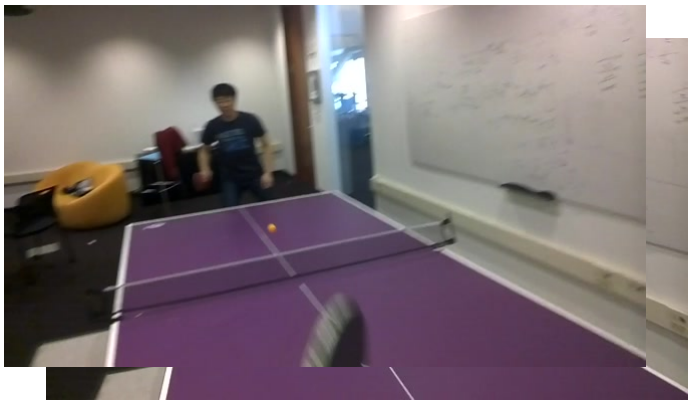
Feed to almost unmodified logic in original software

Example 3: Ping-pong Assistant

- A better chance to win
  - Direct to hit to the left or right based on opponent & ball position
- Not for professionals
- Not for visual impaired

Ping-pong Assistant Workflow

Raw stream (process on pairs)



Symbolic Representation (3-tuple)

<is_playing, ball_pos, opponent_pos>

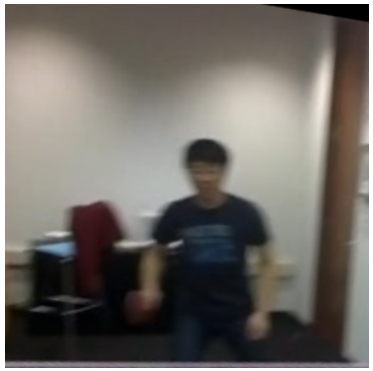
Verbal Feedback

“Left”/ “Right”

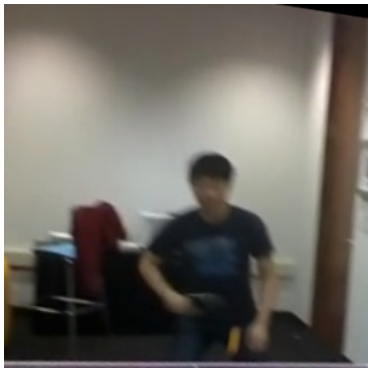
- **Table detection**
- **Opponent detection**
- **Ball detection**

Suggestion based on recent state history

Ping-pong – Opponent Detector



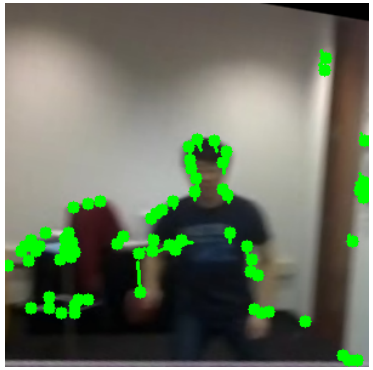
Rotated frame 1



Rotated frame 2

A1: White wall

A2: Dense optical flow



A3: LK optical flow

70 millisecond, error prone

Latency increases by 50%, but more robust

Example 4: Assistance with Crowd-sourced Tutorial

- Deliver context-relevant tutorial videos
  - 87+ million tutorial videos on YouTube
  - State-of-the-art context detector
- E.g. Cooking omelet
  - Recognize egg, butter, etc.
  - Recommend video for same style omelet, using similar tools
- Quickly scale up tasks
- Coarse grained guidance

Tutorial Delivery Workflow

Raw stream (process on video segments)

Symbolic Representation (concept list)

Objects

People

Scene

Action

Video Feedback

- **Dense trajectory feature**
- **State-of-art, slow**
- **1 min processing for a 6 sec video**

- **Indexed 72,000 Youtube videos**
- **Text search using standard language model**

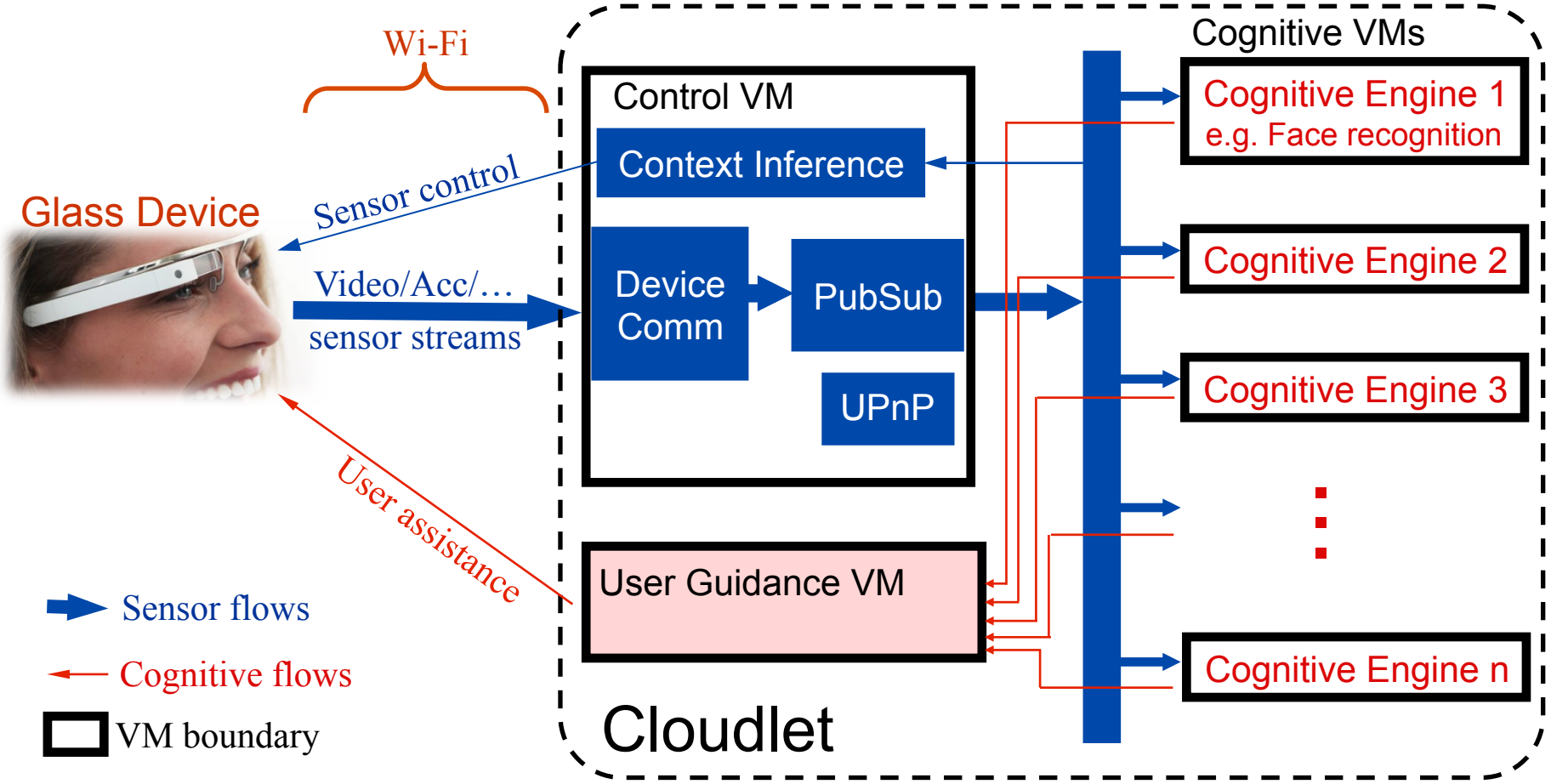
Future Directions

1. Faster Prototyping
2. Improve Runtime Performance
3. Extending Battery Life

Quick Prototyping – State Extractor

Speeding up developing CV algorithms

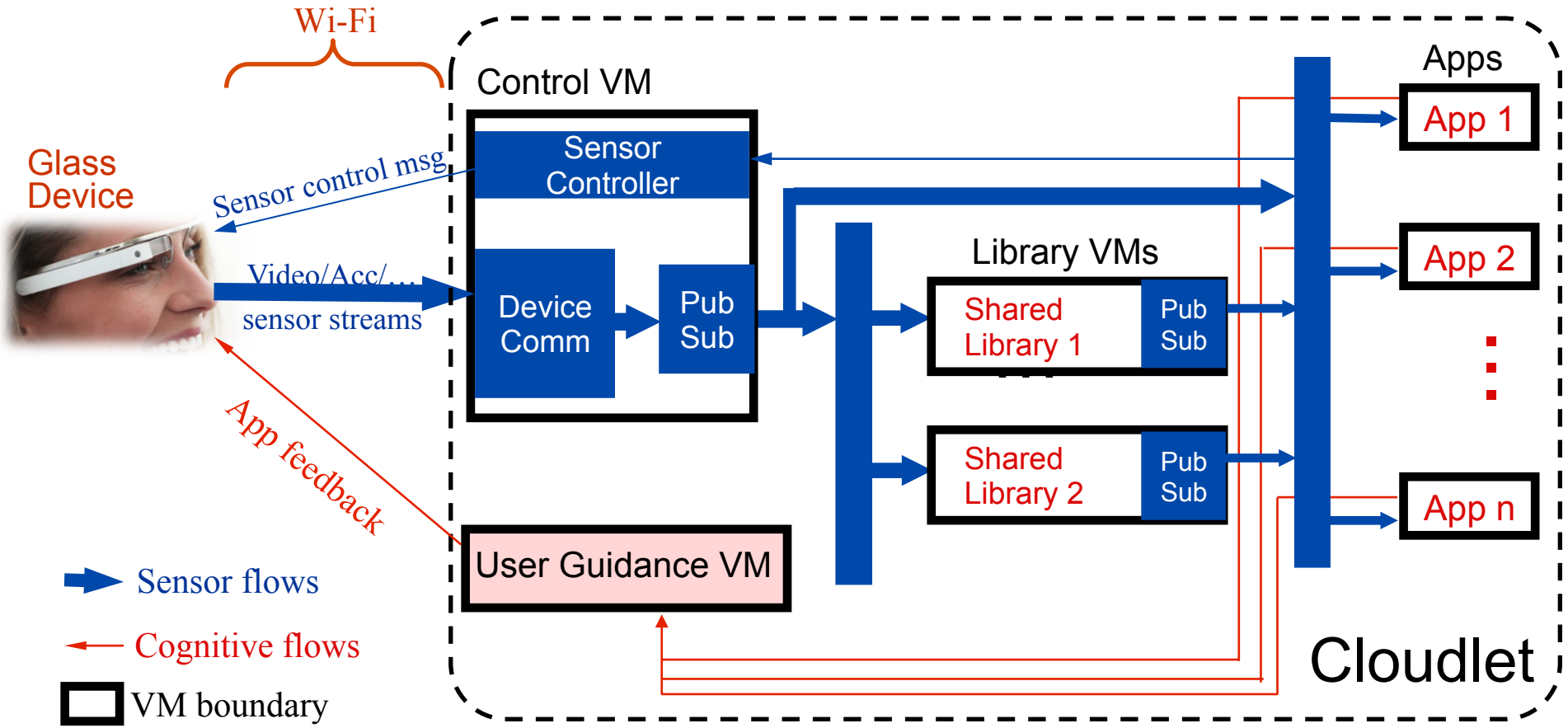
Maybe different applications can share libraries?



Quick Prototyping – State Extractor

Speeding up developing CV algorithms

Maybe different applications can share libraries?



Quick Prototyping – Guidance

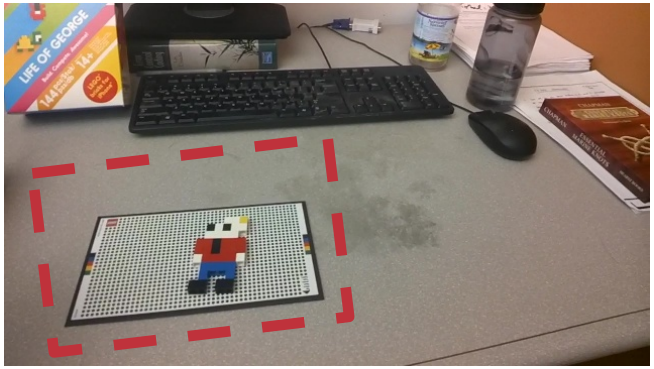
- Easy when state space is small
  - Specify guidance for each state beforehand and match in real-time
  - E.g. Lego, Ping-pong
- Hard when too many states
  - E.g. Drawing, free style Lego
  - "Guidance by example": learn from crowed-sourced experts doing the task

Improving Runtime Performance

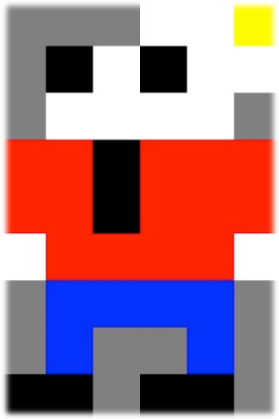
- Leverage multiple algorithms
  - Do exist. Maybe just different parameters
  - *Tradeoff* between accuracy and speed
    E.g. Ping pong opponent detector
- Accuracy of an algorithm depends on
  - Lighting, background
  - User, user's state
  - Won't change quickly within a task
- Run all, use optimal!

Extending Battery Life

- Region of Interest (ROI) exists for some tasks
  - Lego (board)
  - Drawing (paper)



- ROI doesn't move quickly among frames
  - Cheap computation on client
  - Transmit only potential ROI



Early Implementation Experience with Wearable Cognitive Assistance Applications

Zhuo Chen, Lu Jiang, Wenlu Hu, Kiryong Ha, Brandon Amos,
Padmanabhan Pillai, Alex Hauptmann, and Mahadev Satyanarayanan

Carnegie Mellon University

intel

Backup Slides

Glass-based Virtual Instructor

1. Understand user's state
   - Real instructor: use eyes, ears, and knowledge
   - Virtual: sensors, computer vision & speech analysis + task representation
2. Provide guidance to user
   - Real instructor: speak, or show demos
   - Virtual: text/speech/image/video

An Example

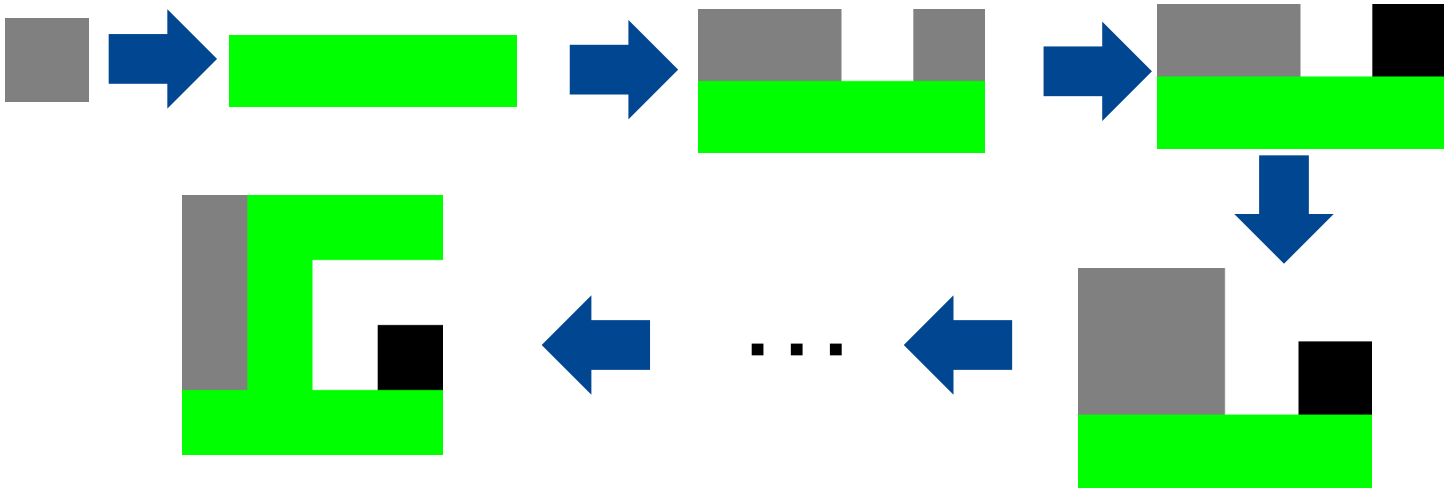
- Making Butterscotch Pudding
  - Glass gives guidance (e.g. step-by-step instructions)
    - E.g. “Gradually whisk in 1 cup of cream until smooth”
  - Glass checks if user is doing well
    - E.g. Cream amount ok? Smooth enough?
  - Guidance adjusted based on user’s progress
    - E.g. “Next step is …” **OR** “Add more cream!”

Task Representation

- Matrix representation of Lego state

```
[[0, 2, 2, 2],
 [0, 2, 1, 1],
 [0, 2, 1, 6],
 [2, 2, 2, 2]]
```

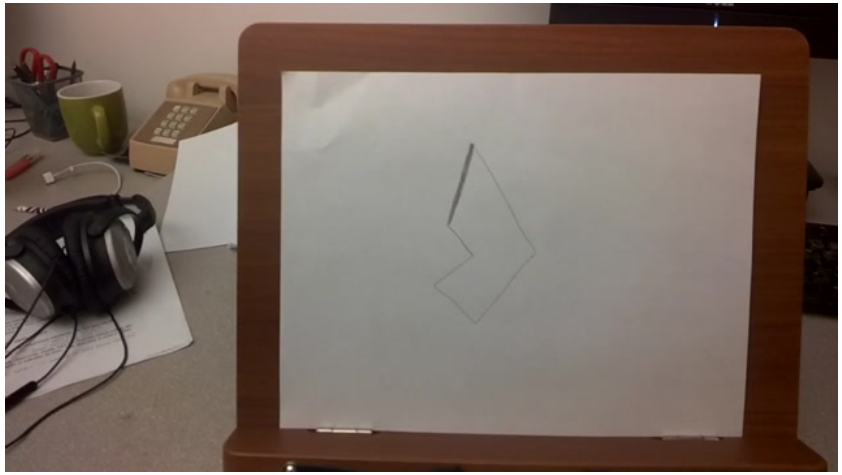
- Task represented as a list of states

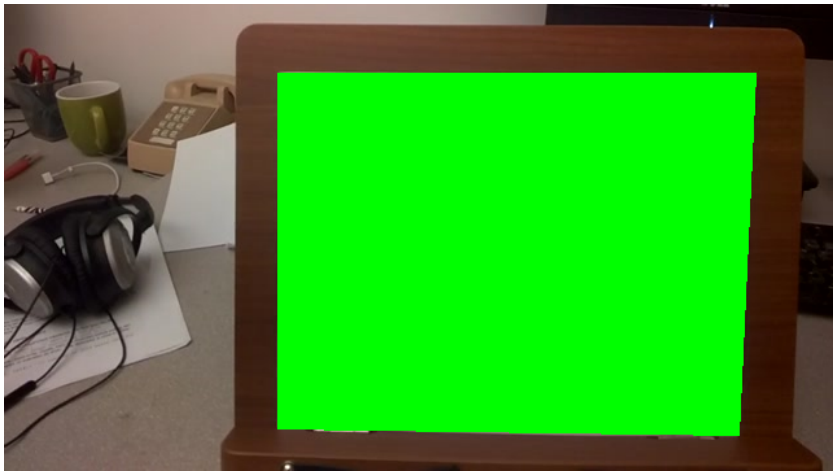


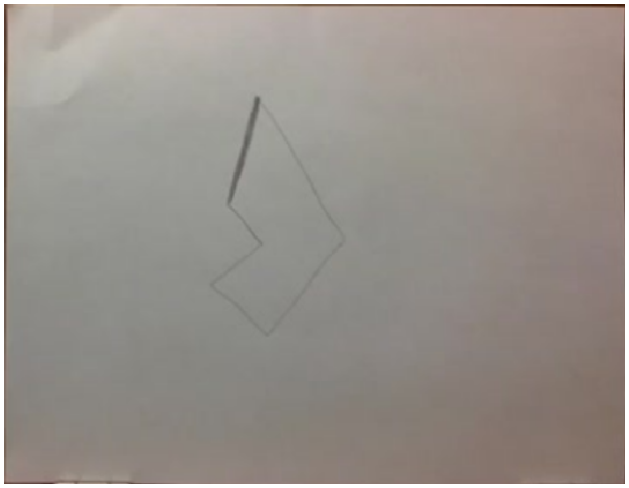
Guidance to the User

- Speech guidance
  - "Now find a 1x4 piece and add it to the top right of the current model"
  - " This is incorrect. Now move the… to the left"
- Visual guidance
  - Animations to show the three actions
- Demo

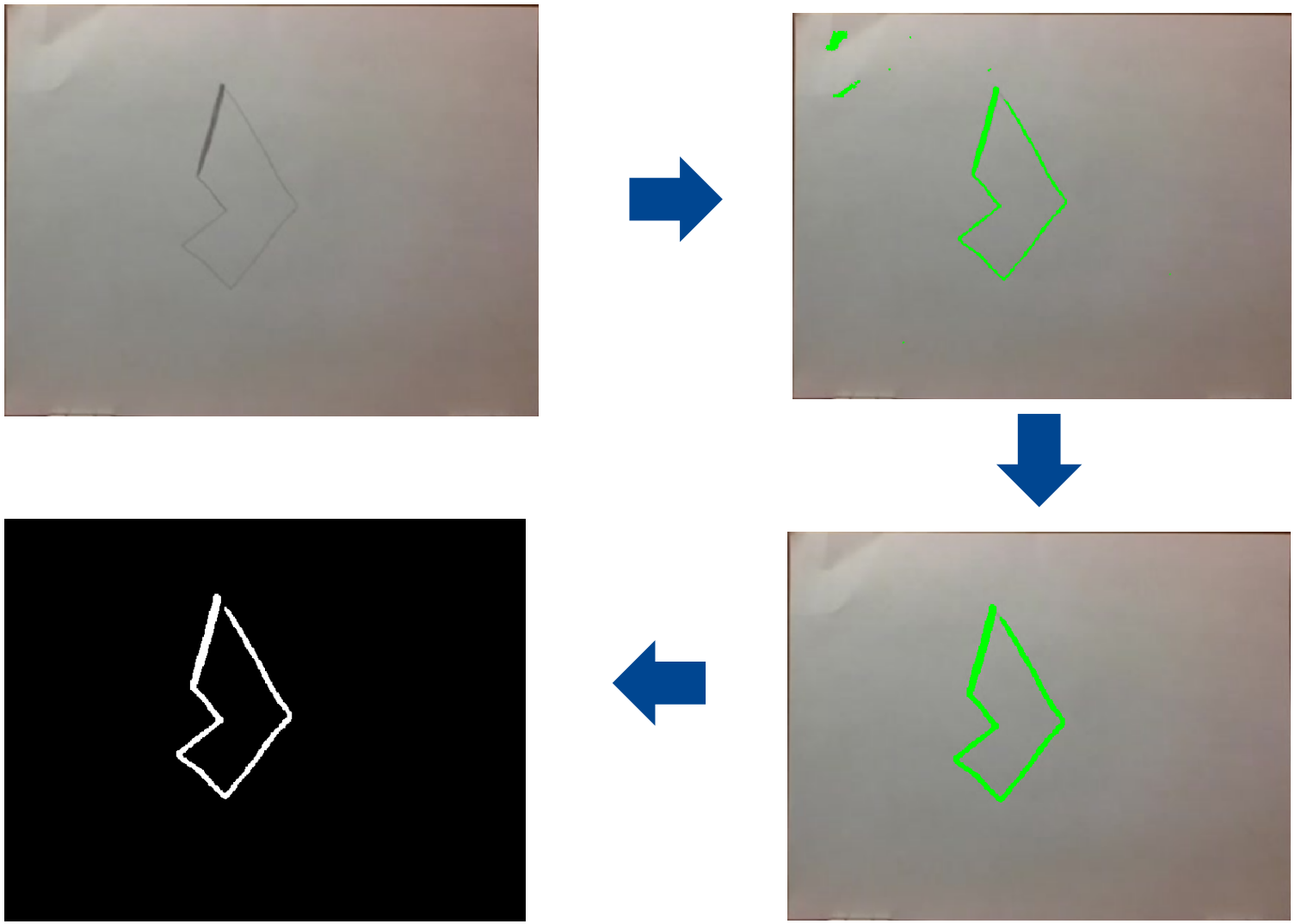
State Extractor (1)





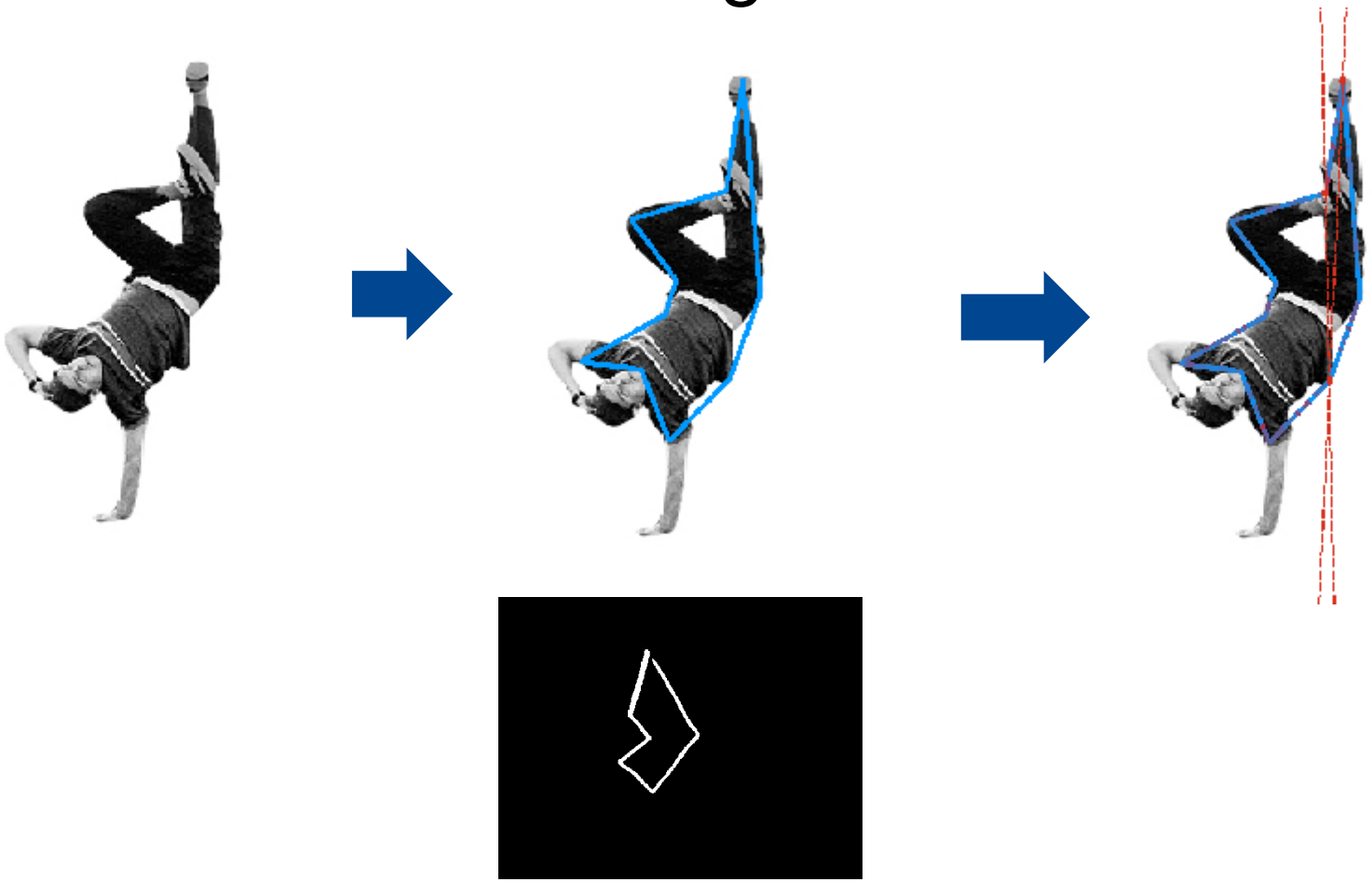


State Extractor (2)



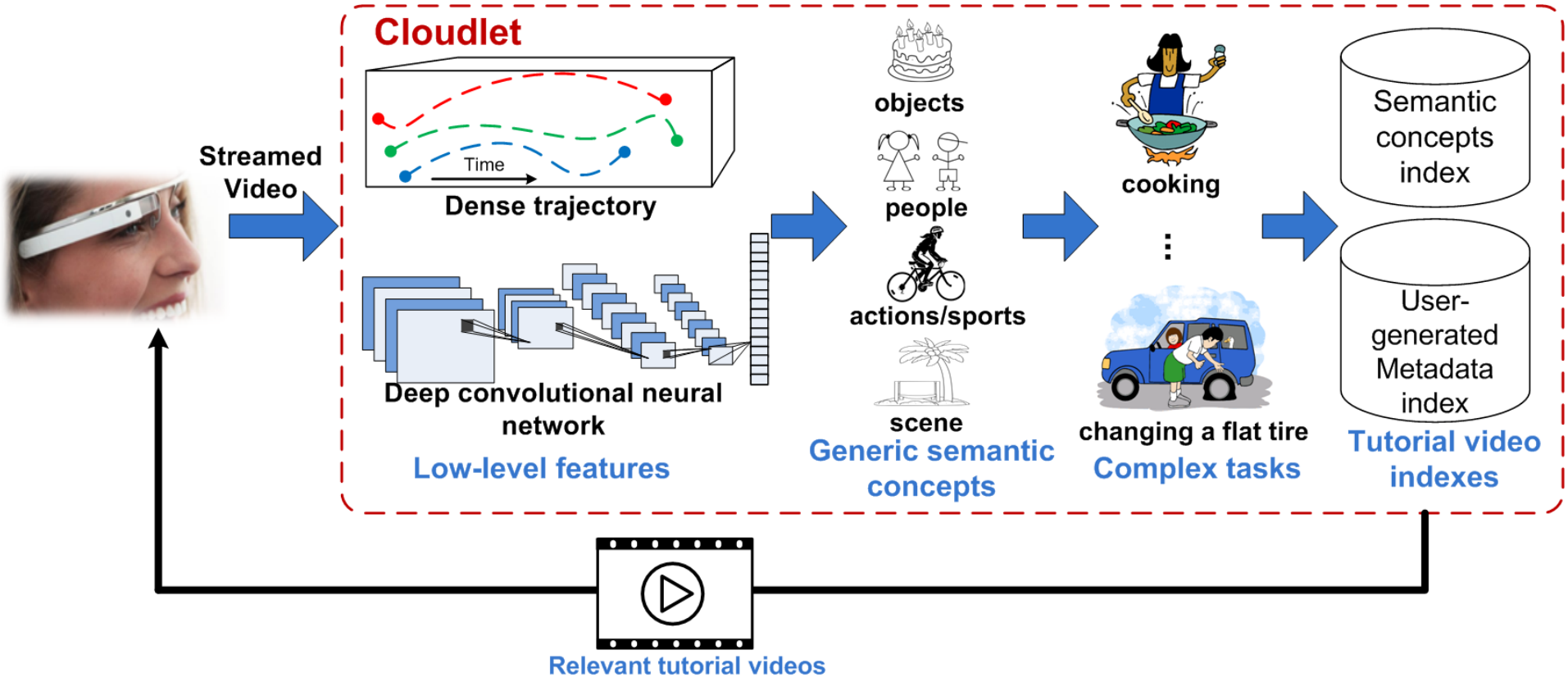
Guidance

- Call function in original software



State Extractor – Table Detection

State Extractor



1 min to detect context from a 6 sec video

Symbolic Representation

- Concept list + high level task
  - Can detect 3000+ concepts

Semantic Category	Examples
People	male, baby, teenager, girl, 3 or more people
Scene	beach, urban scene, outdoor, forest
Object	car, table, dog, cat, guitar, computer screen
Action	shaking hand, cooking, sitting down, dancing
Sports	cycling, skiing, bullfighting, hiking, tennis

Quick Prototyping – Guidance

- Hard when too many states
- Learn from examples
  - Record the task performance from multiple (maybe crowd-sourced) users
  - Run state extractor to extract state chain
  - For a new state from current user, find the optimal match to provide guidance
  - Performance improved as more people use

Framework for Easy Parallelism

- Inter-frame parallelism
  - Easier
  - Improve throughput, not latency
- Intra-frame parallelism
  - Scanning window – detection based on recognition
  - Extract local features from a big picture
  - Sprout

Identify State Change

- Full processing needed only for new state
- Savings can be huge!
  - For a 2 minute Lego task, there are only 10 states
  - Only 10 images need to be transmitted! (not 1800!)
  - The question is which 10…
- Use “cheap” sensor to detect state change
  - Turn “expensive” sensor on when there is

Identify State Change

- Instructor doesn't watch you all the time!
  - Probably just after giving some guidance, she won't watch you.
    - After guidance, turn camera off for some time.
  - Instructor has time expectation of each step
    - Can set expectation learned from other users
    - Adapt to the current user
  - Instructor will check regularly
    - Transmit image at very low sampling rate
    - Turn accelerometer on after some time