Blind Person Assistance using Object Detection

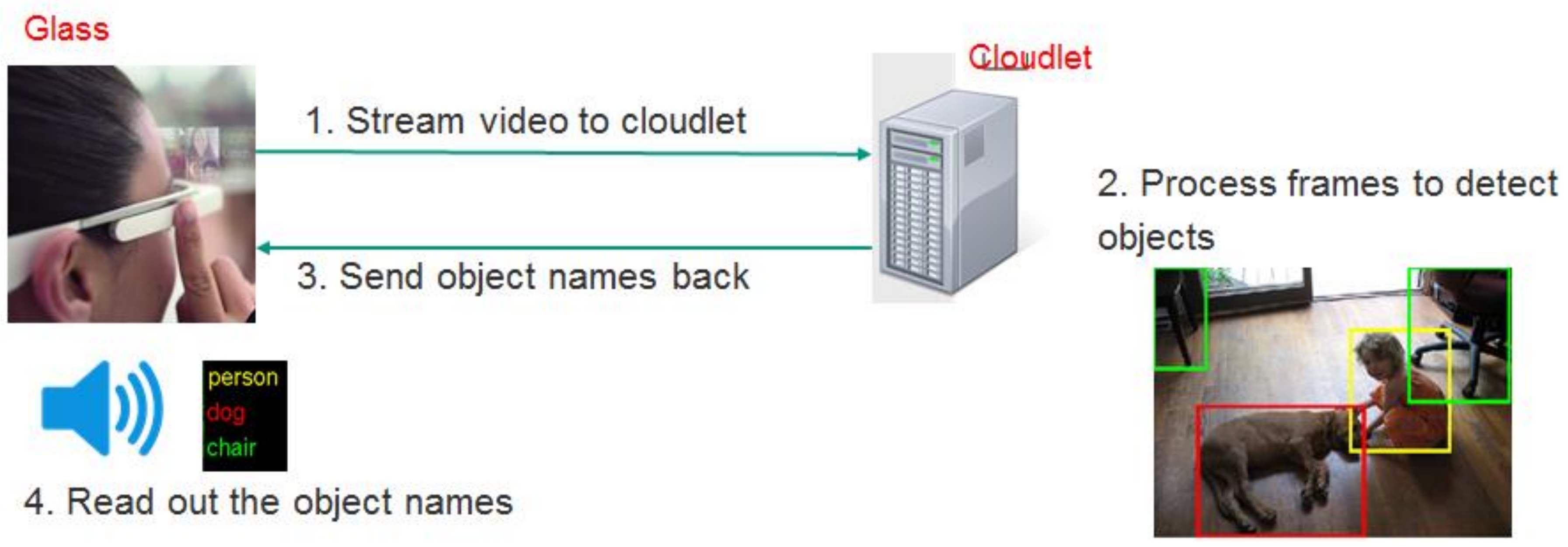
Shruti Dhoot & Vrushali Bhutada

Mentor: Brandon Amos

Objective

Assist visually impaired Google Glass users by providing spoken cues of objects they are looking at

Methodology



Glass

Cloudlet

1. Stream video to cloudlet

2. Process frames to detect objects

3. Send object names back

4. Read out the object names

Combines the first-person image capture and sensing capabilities of Glass with remote processing to perform near real-time object detection

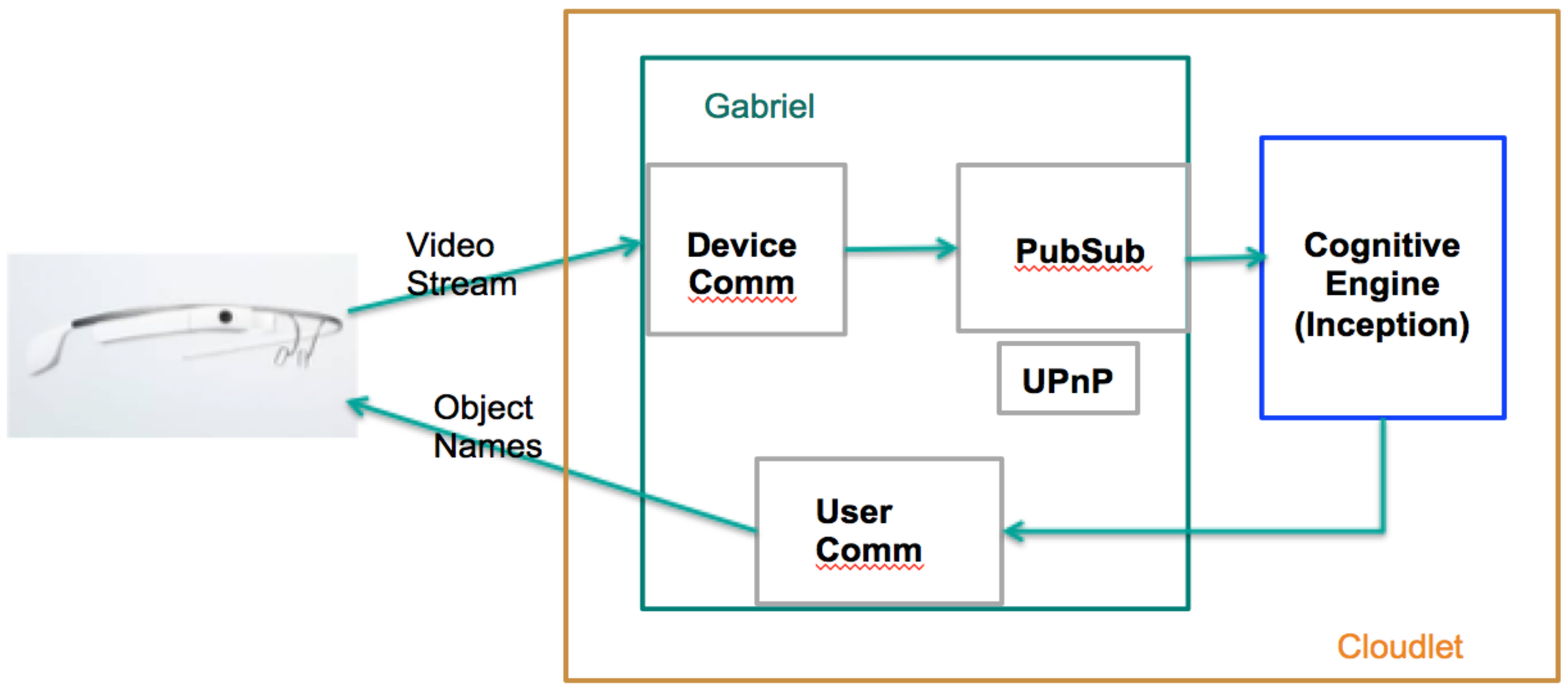
Key Components

Cloudlet

- *Data center in a box*
- Represents the middle tier of a 3-tier hierarchy: mobile device - cloudlet - cloud
- Provides low end-to-end latency and high bandwidth

Gabriel

- Offloads compute-intensive operations from Glass to Cloudlet
- Gracefully degrades in case of network failures
- Written in Python and Java (Android Client)



Video Stream

Object Names

Gabriel

Device Comm

PubSub

UPnP

Cognitive Engine (Inception)

User Comm

Cloudlet

Inception

- Deep neural network architecture
- Achieves the new state of the art for object detection and classification
- Efficient in terms of power and memory use (not just accuracy)
- Can be used on large datasets at reasonable cost
- Treated it as a black box where we would input a video frame and top 5 predictions were returned
- Written in Python - so was easier to integrate with Gabriel

Object Detection Models

Inception	Teradeep
Trained on 200 classes of objects	Trained on 1000 classes of objects
Written in Python (using Caffe)	Written in Lua (using Torch)
Slower as it was developed for image classification. Need to write each video frame to a file to process it.	Faster as it was developed to process camera feed. Provides near real-time object detection.
Research project (source code available)	Available as a commercial product (Object detection source code is available)

Results

Processing time of video frames ranged from 200-300 ms per frame. Predictions were more accurate when the objects belonged to the same categories as the training data set.

Interview

We interviewed a blind person to understand his needs better.

Key insights -

- Need assistance in unfamiliar environments
- Would like to know where useful items are
- Can manually control app via speech commands
- Should detect signboards like Exit, Elevator, etc
- Would help people who have newly lost vision

Concerns -

- Might violate other people’s privacy if video is streamed continuously

Limitations

- Cues are provided at slower rate than the response rate
- Inception requires video frames to be written to files
- Inception can classify only 200 categories of objects accurately

Future Work

- Adopt a model which can detect more classes of objects
- Detect objects intelligently based on user’s preferences
- Allow tuning of rate at which spoken cues are provided
- Provide directions to an object detected by Glass

Takeaways

- Can perform compute intensive task like image processing using a resource poor mobile device
- Can detect specific objects accurately
- Can potentially eliminate stress and guesswork in new places

References

1. *Towards Wearable Cognitive Assistance* - Kiryong Ha, et al
2. *Going Deeper with Convolutions* - Christian Szegedy, et al
3. http://elijah.cs.cmu.edu/
4. https://github.com/teradeep/demo-apps
5. https://github.com/google/inception