AiR

Artificial Intelligence for Robots

Development of Human-Care Robot Technology for Aging Society

2019.10.14

@SHRI Workshop / RO-MAN 2019

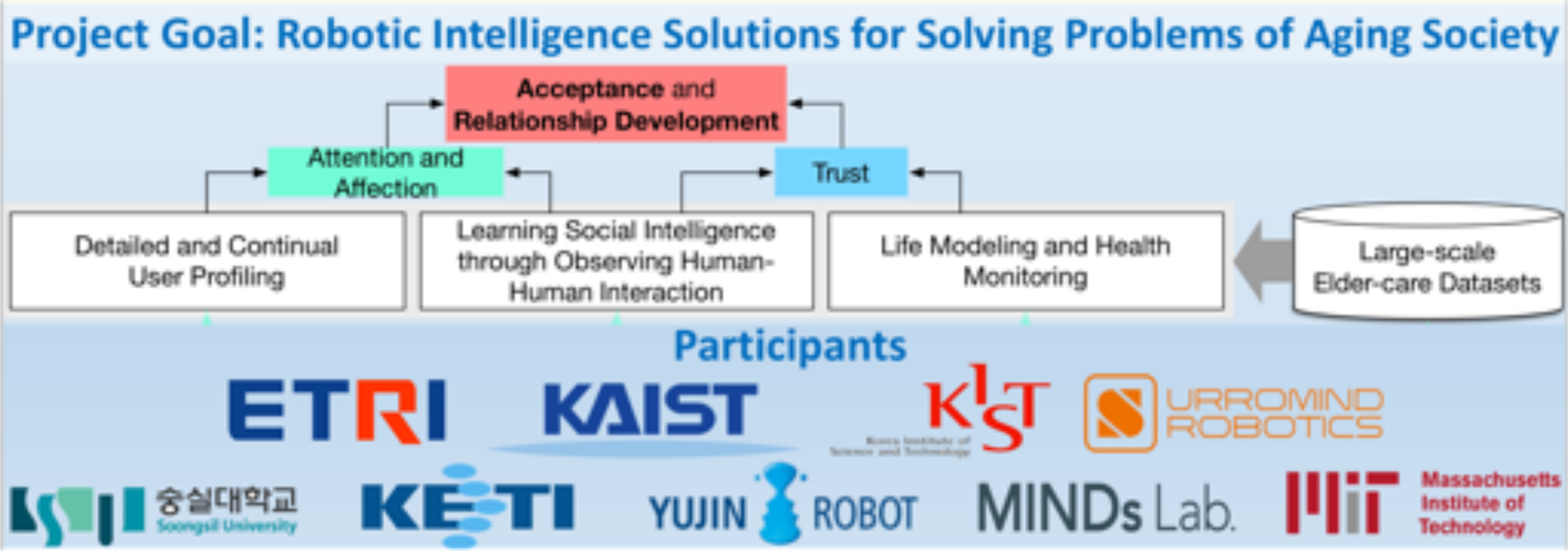
Minsu Jang

HRI Lab, ETRI

ETRI KIST Korea Institute of Science and Technology KETI KAIST

숭실대학교 Soongsil University URROMIND ROBOTICS YUJIN ROBOT MINDs Lab. MIT Massachusetts Institute of Technology

Research Goal



Project Goal: Robotic Intelligence Solutions for Solving Problems of Aging Society

Acceptance and Relationship Development

Attention and Affection

Trust

Detailed and Continual User Profiling

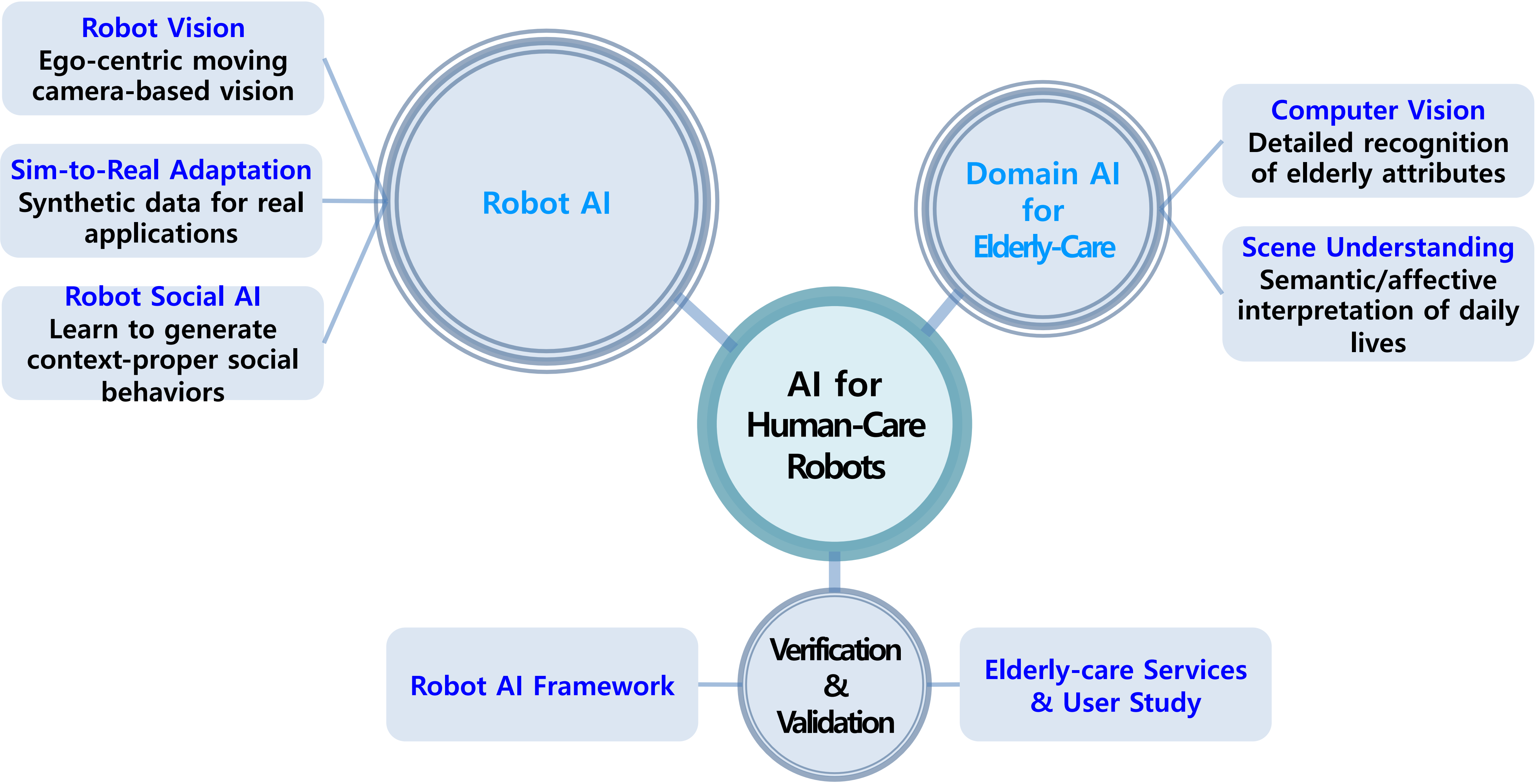
Learning Social Intelligence through Observing Human-Human Interaction

Life Modeling and Health Monitoring

Large-scale Elder-care Datasets

Participants

Research Issues



AI for Human-Care Robots

Robot AI

Robot Vision
Ego-centric moving camera-based vision

Sim-to-Real Adaptation
Synthetic data for real applications

Robot Social AI
Learn to generate context-proper social behaviors

Domain AI for Elderly-Care

Computer Vision
Detailed recognition of elderly attributes

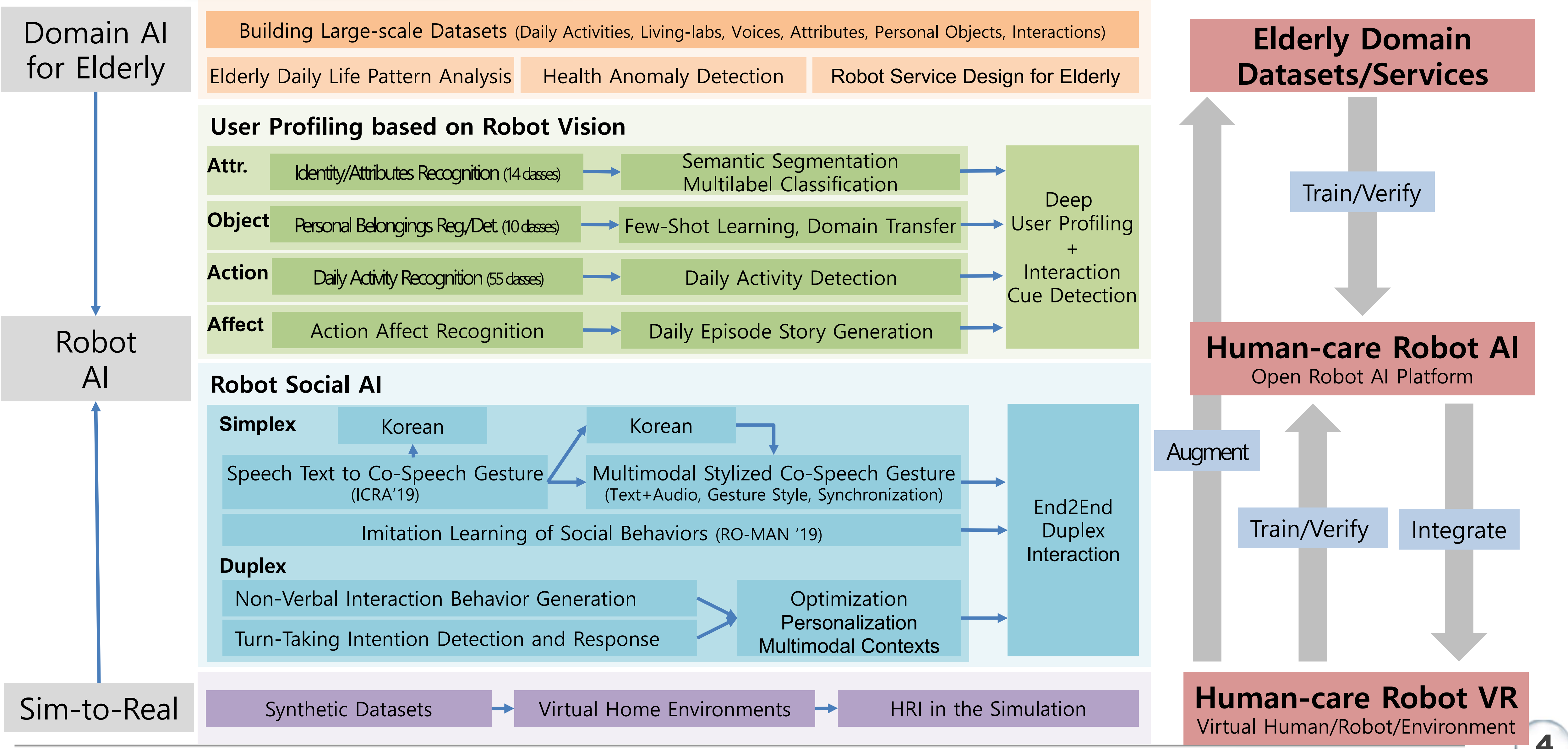
Scene Understanding
Semantic/affective interpretation of daily lives

Verification & Validation

Robot AI Framework

Elderly-care Services & User Study

Research Roadmap



Domain AI for Elderly

Building Large-scale Datasets (Daily Activities, Living-labs, Voices, Attributes, Personal Objects, Interactions)

Elderly Daily Life Pattern Analysis

Health Anomaly Detection

Robot Service Design for Elderly

Robot AI

User Profiling based on Robot Vision

Attr. Identity/Attributes Recognition (14 classes) → Semantic Segmentation Multilabel Classification →

Object Personal Belongings Reg./Det. (10 classes) → Few-Shot Learning, Domain Transfer →

Action Daily Activity Recognition (55 classes) → Daily Activity Detection →

Affect Action Affect Recognition → Daily Episode Story Generation →

Deep User Profiling + Interaction Cue Detection

Robot Social AI

Simplex

Korean

Speech Text to Co-Speech Gesture (ICRA'19)

Korean

Multimodal Stylized Co-Speech Gesture (Text+Audio, Gesture Style, Synchronization) →

Imitation Learning of Social Behaviors (RO-MAN '19) →

Duplex

Non-Verbal Interaction Behavior Generation

Turn-Taking Intention Detection and Response

Optimization Personalization Multimodal Contexts →

End2End Duplex Interaction

Sim-to-Real

Synthetic Datasets → Virtual Home Environments → HRI in the Simulation

Elderly Domain Datasets/Services

Train/Verify

Human-care Robot AI
Open Robot AI Platform

Augment

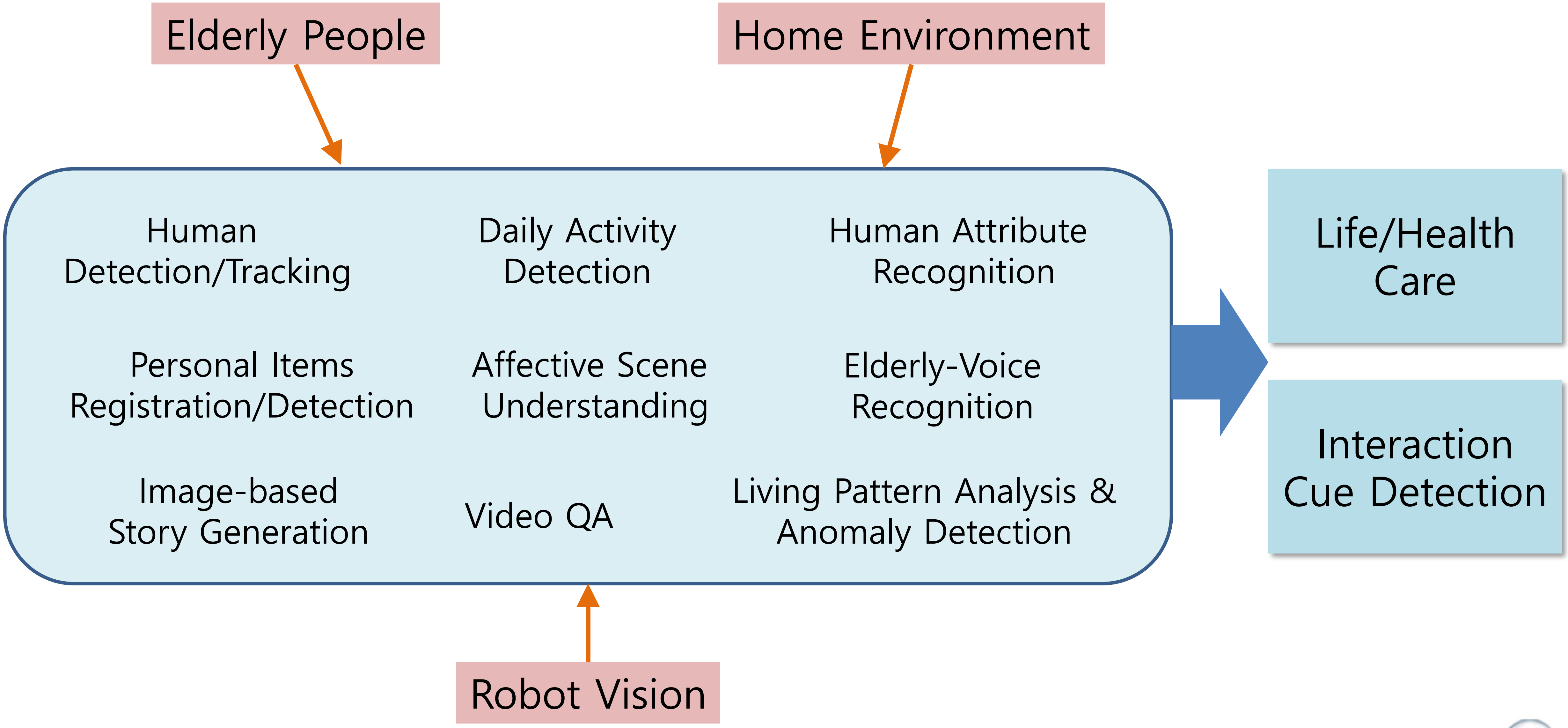
Train/Verify

Integrate

Human-care Robot VR
Virtual Human/Robot/Environment

Domain AI
for Elderly-Care

Domain AI for Elderly-Care



Elderly People

Home Environment

Human Detection/Tracking

Daily Activity Detection

Human Attribute Recognition

Personal Items Registration/Detection

Affective Scene Understanding

Elderly-Voice Recognition

Image-based Story Generation

Video QA

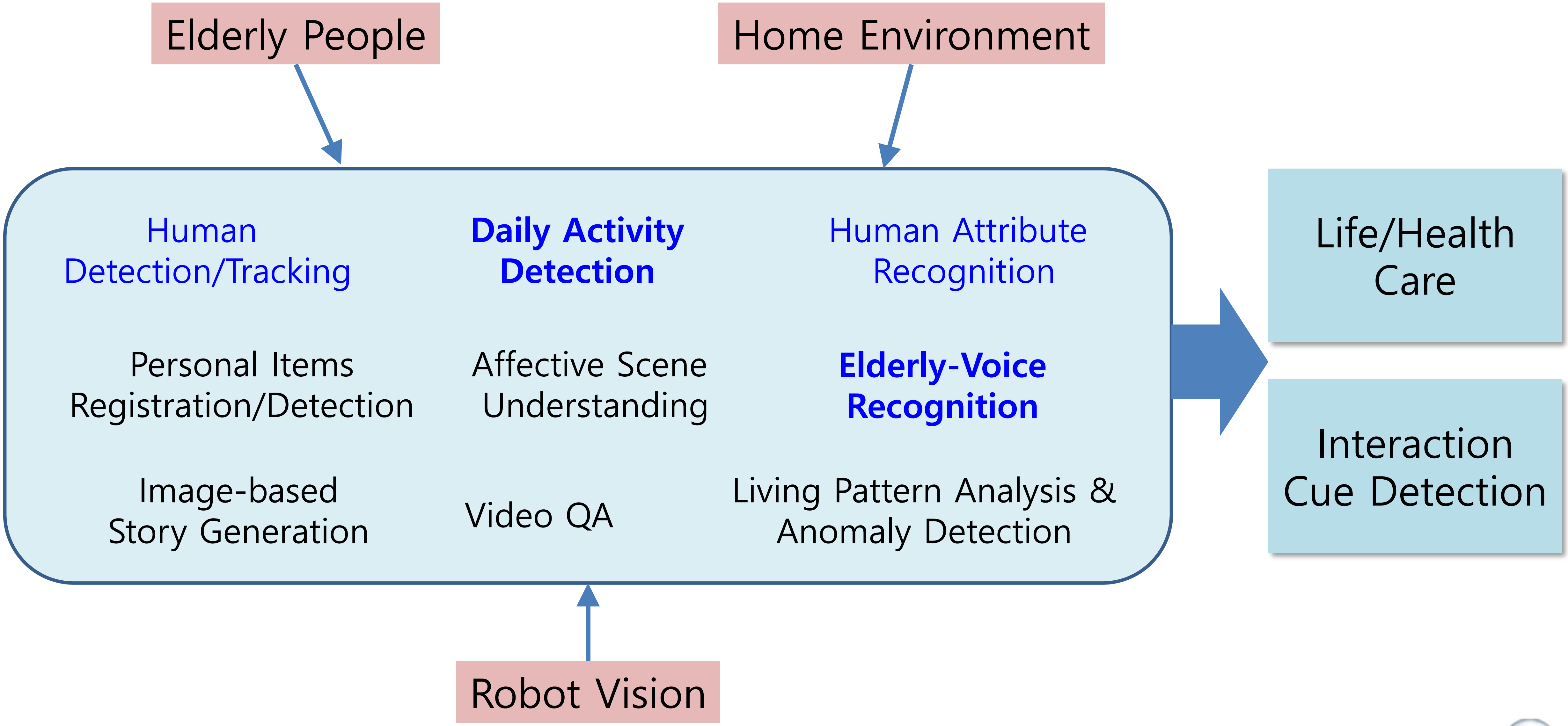
Living Pattern Analysis & Anomaly Detection

Robot Vision

Life/Health Care

Interaction Cue Detection

Domain AI for Elderly-Care



Elderly People

Home Environment

Human Detection/Tracking

Daily Activity Detection

Human Attribute Recognition

Personal Items Registration/Detection

Affective Scene Understanding

Elderly-Voice Recognition

Image-based Story Generation

Video QA

Living Pattern Analysis & Anomaly Detection

Life/Health Care

Interaction Cue Detection

Robot Vision

Domain AI for Elderly-Care

- Hypothesis: X of elderly people are very different from X of young adults. (X=motion, fashion, verbal features, facial expressions etc.

We need data from elderly people.

Daily Activity Detection for Elderly People

Activity Selection

Method	**Goal**	Select most frequent activities of older people
	How	Observing one day of older people
	Participants	53 Elderly People (age > 65)
	Dates	2017-06-15 ~ 2017-07-05
Result	**No. observed activities**	245
	Frequent activities	1. Watching TV 2. Meal-related activities (eating, preparing foods, washing dishes) 3. Defecation (using toilet) 4. Phone call 5. Taking medications 6. Washing face and brushing teeth 7. Wearing and taking off clothes
	Frequently used objects	Mobile phone, Remote, Eyeglasses, Beds, Medicine, Cups

Daily Activity Detection for Elderly People

Activity Selection

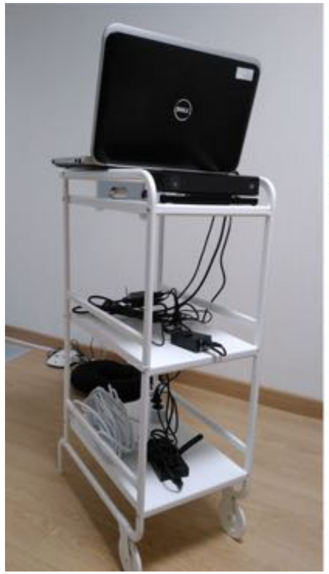
- We selected 55 frequent activities for detection.
- Selected Activities:

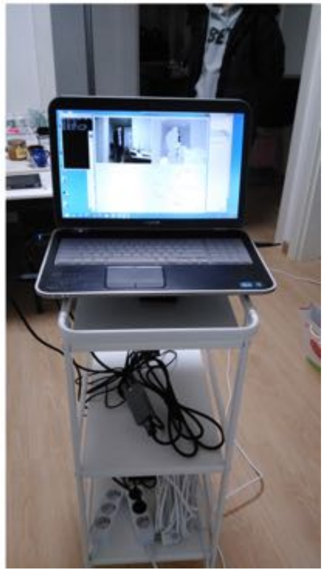
  see table

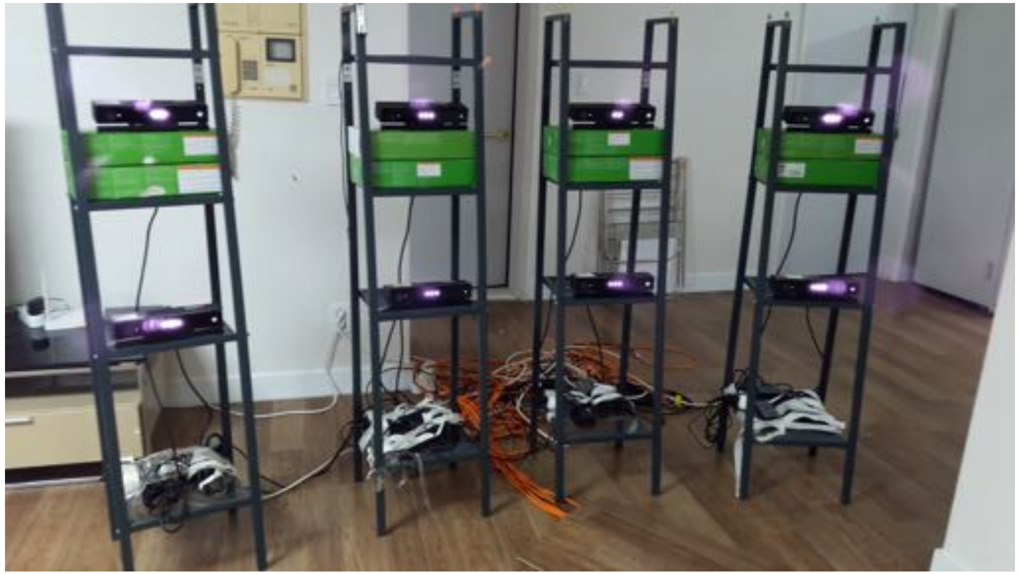
Daily Activity Detection for Elderly People

Data Acquisition: Considerations

- Real-World Data: Testbeds, Living Labs
- Multi-Modality: RGB-DS
- Multiple Views: 8 different camera positions
- Moving Camera







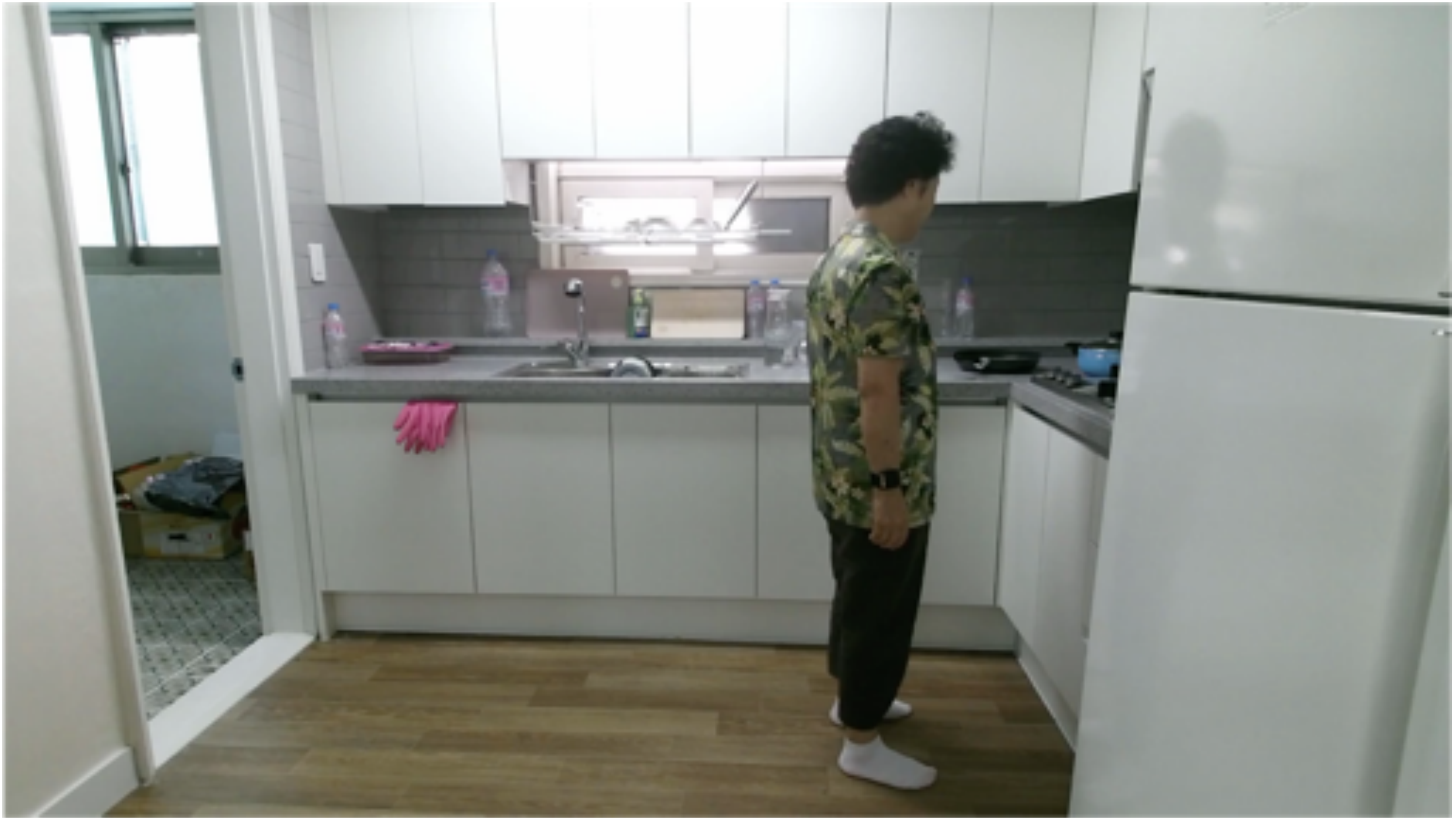
Daily Activity Detection for Elderly People

Data Acquisition: Environments and Participants

- Living Labs: homes where elderly people actually are living
  - Real life situations without intervention (slight interventions are being tried though)
  - Moving camera using a cart operated by a human operator
- Testbed: An apartment house for data collection and experiments
  - 55 activities are acted by participants
  - RGB-D cameras in 8 different viewpoints

Daily Activity Detection for Elderly People

Data Acquisition: Testbed



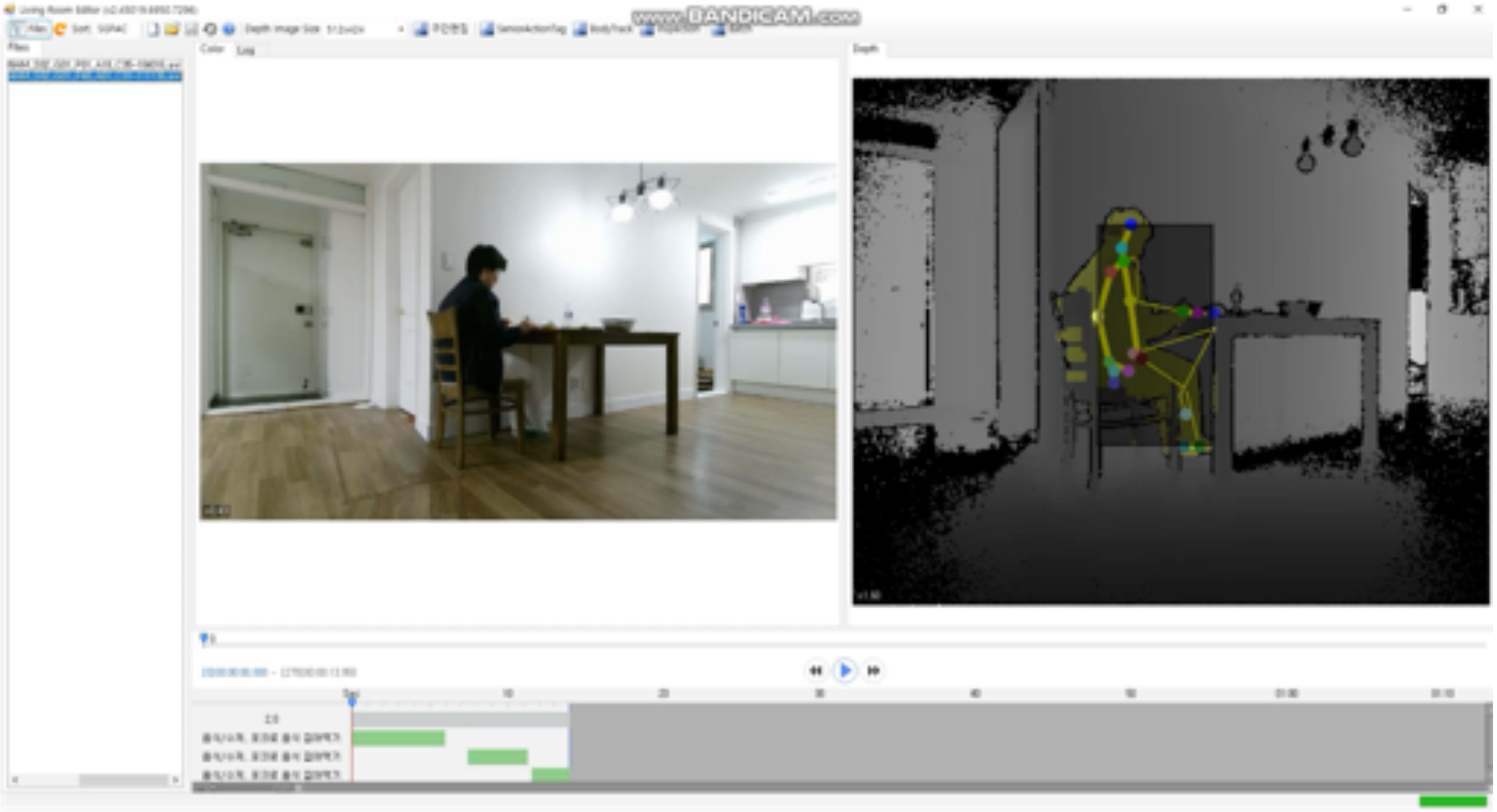
Daily Activity Detection for Elderly People

Data Acquisition: Living Labs

Daily Activity Detection for Elderly People

Data Acquisition: Annotations & Validations

- 3D Skeleton Joints, Activity Endpoints



Daily Activity Detection for Elderly People

Data Acquisition: Elderly Activity Datasets

- Data Format: RGB / Depth / Skeleton
- Living Labs
  - Participants: 18 homes (2017 ~ present)
  - 200 hours of 6,048 video clips
- Testbed
  - Participants: 50 elderly people / 50 young adults
  - 111,672 sets of video data

To be publicly available before in 2020
http://ai4robot.github.io

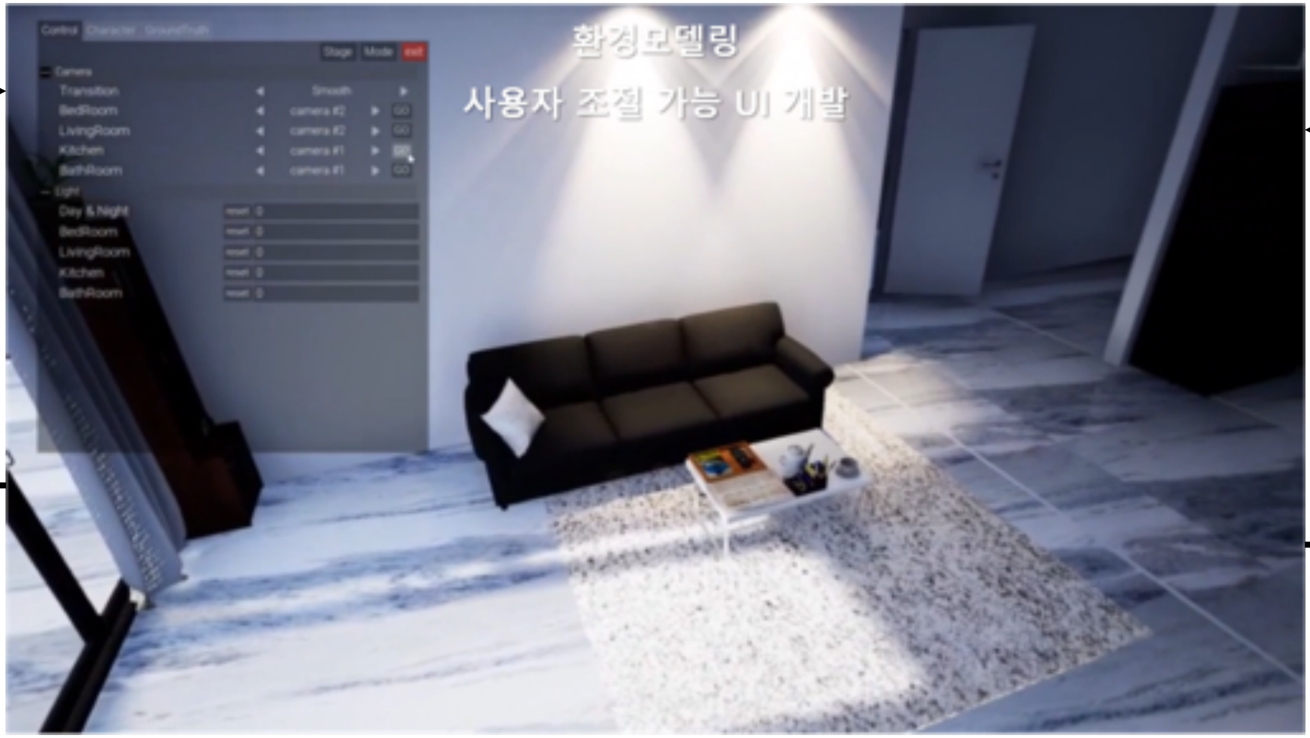
Daily Activity Detection for Elderly People

Synthetic Data Generation

Virtual Home Robot Environment

Parameter Variations →

← Large-scale Synthetic Human, Activity and Environment Data



← Robot Model Demonstrations Interaction

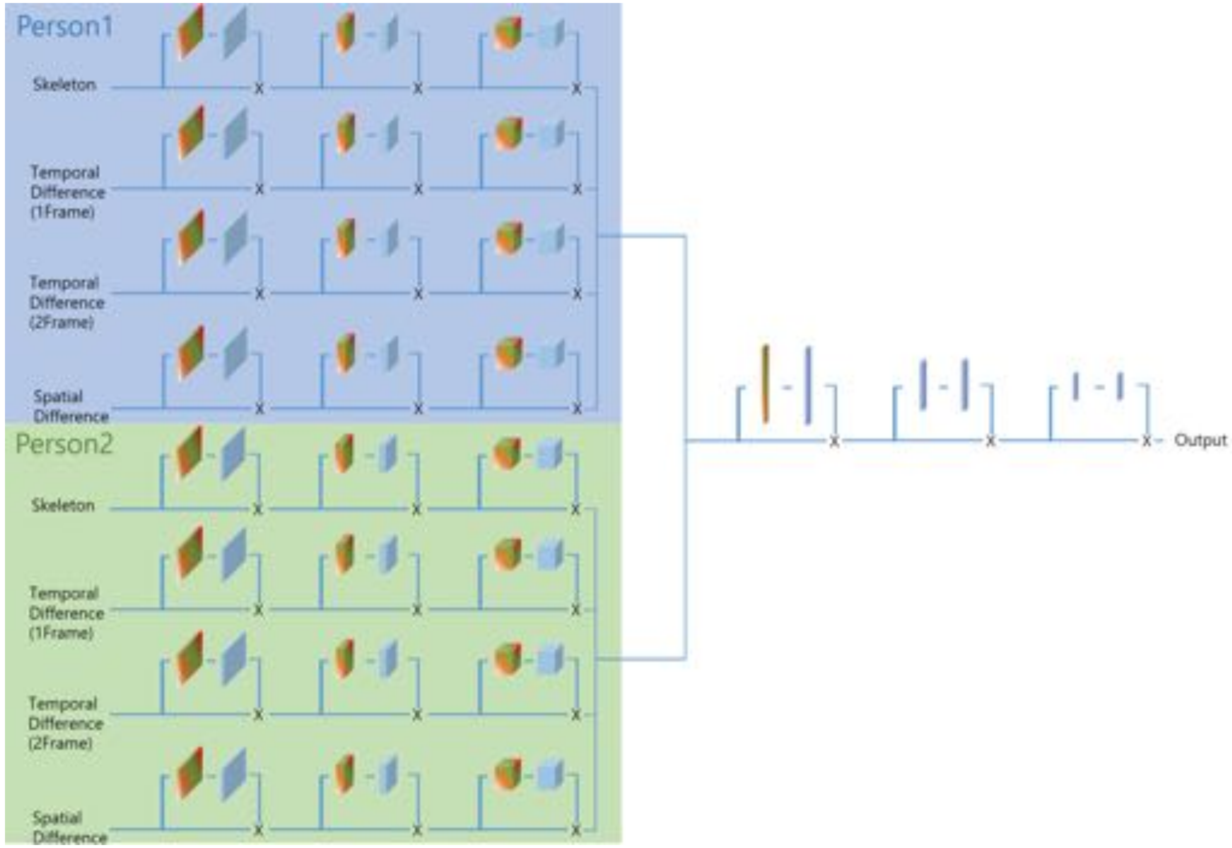
→ Robot AI Trained

To be publicly available in 2020
http://ai4robot.github.io

Daily Activity Detection for Elderly People

Activity Detection

- Trainable Activation-based RNN


Person1
Skeleton
Temporal Difference (1Frame)
Temporal Difference (2Frame)
Spatial Difference
Person2
Skeleton
Temporal Difference (1Frame)
Temporal Difference (2Frame)
Spatial Difference
Output

Benchmark with NTU dataset

Model	Org	Performance	Data
TS-CNN	Ludwig Maximilian University	83.2%	S
C-ConvNet	Univ. of Wollongong	86.4%	RGBD
HCN	Hikvision Research Institute	86.5%	S
Glimpse Clouds	Univ. Lyon & INRIA	86.6%	RGB
I3D	DeepMind	88.6%	D
SLnL-rFA	Chinese Academy of Sciences	89.1%	S
I3D	DeepMind	89.5%	RGB
Evolution of Pose Estimation Map	Paris Seine University	91.7%	RGBS + Heatmap
Ours	**ETRI**	**90.4%**	**RGBS**

Jang, Jinhyeok et al., “Deep Asymmetric Networks with a Set of Node-wise Variant Activation Functions.”, arXiv preprint arXiv:1809.03721 (2018)

Daily Activity Detection for Elderly People

Activity Detection

- Hypothesis Validation
  "Is it plausible that activity patterns of elderly people are very different from those of young adults?" "Yes, maybe…"

	Tested with elderly data	Tested with young data
Trained with elderly data	87.69	68.99
Trained with young data	74.87	85.00
Trained with mixed data	84.78	82.05

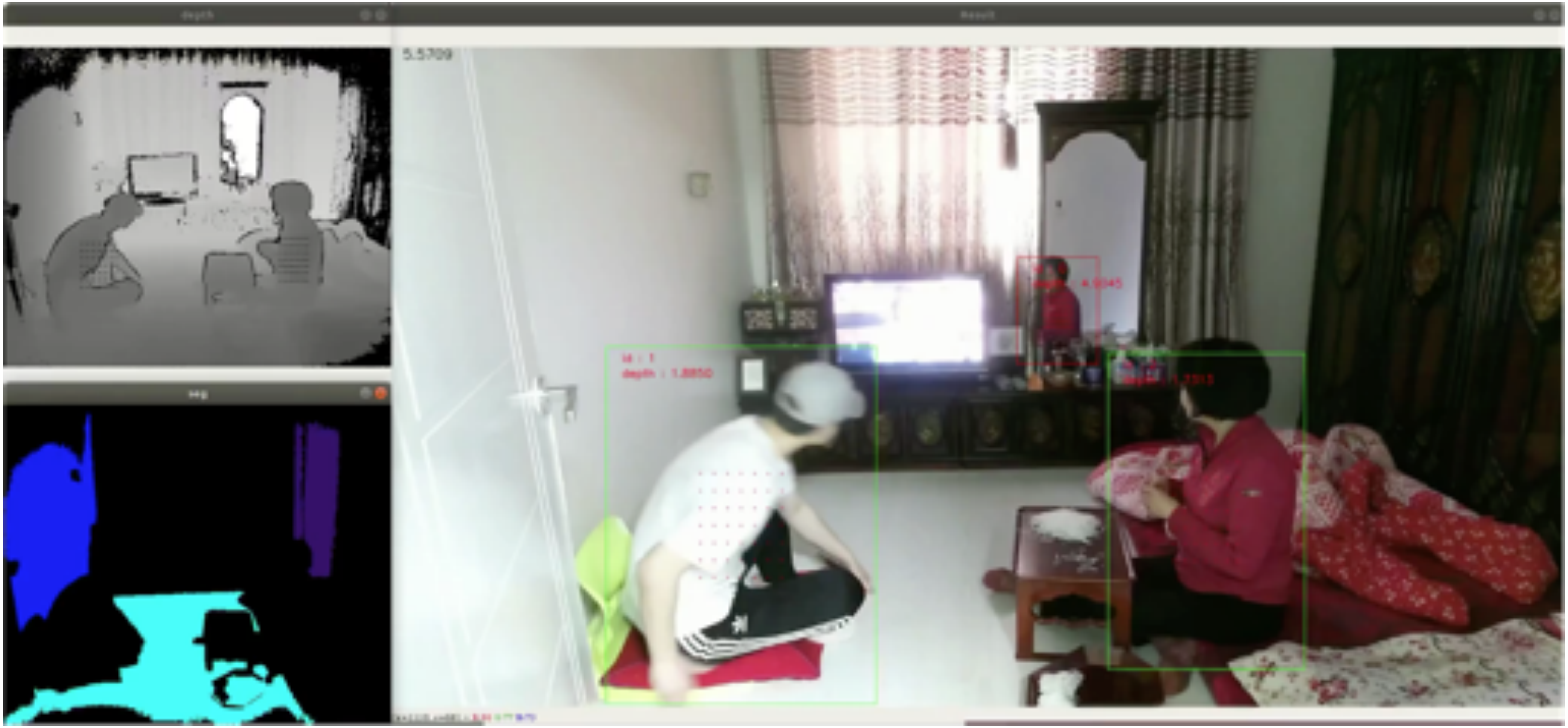
Human Detection/Tracking

Issues

- Robot Vision: Moving Camera
- Home Environment: Cluttered, Partial Body Exposure
- Reflections on the mirrors, reflective planes
- Robust Re-identification

Human Detection/Tracking

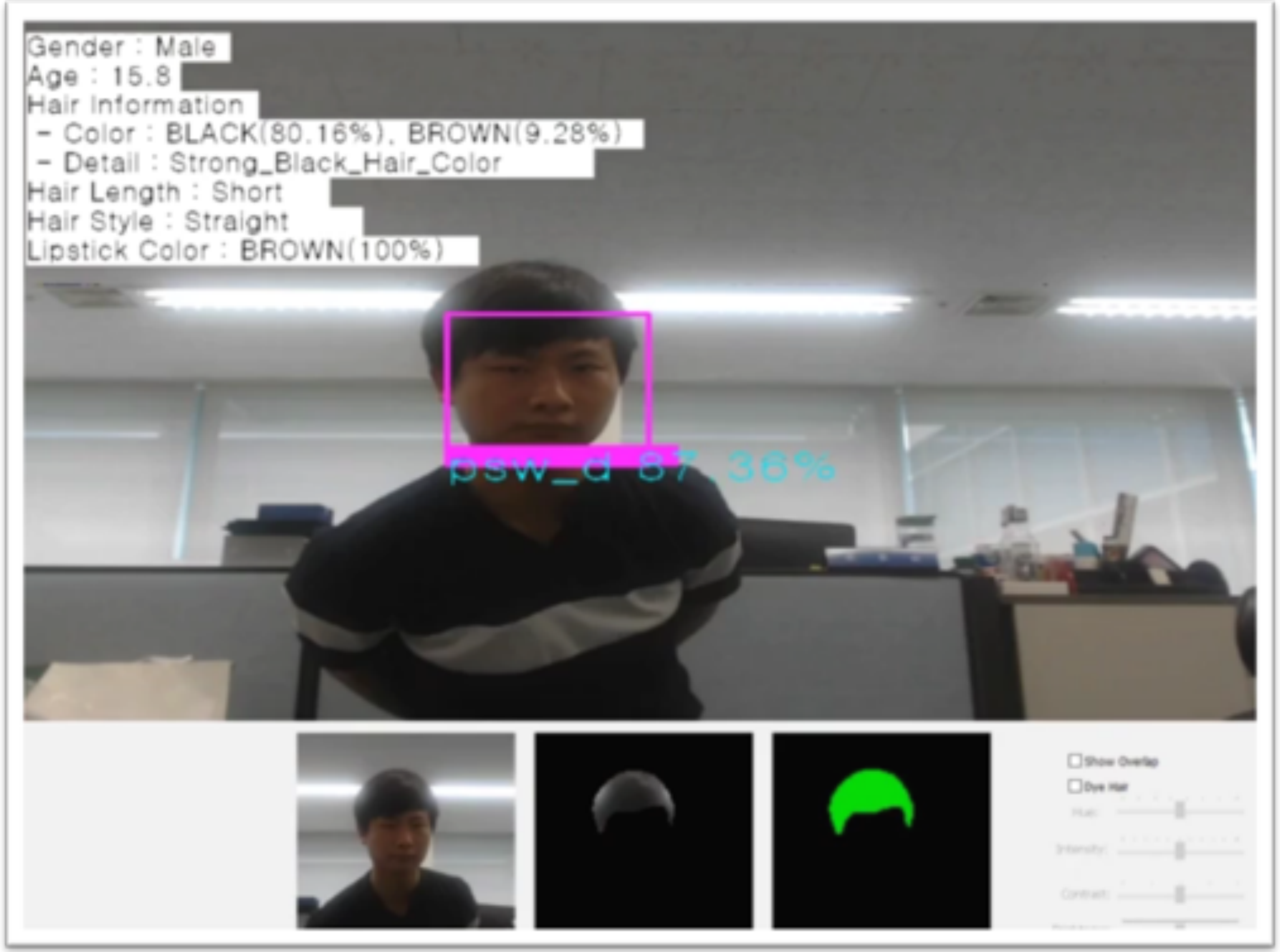
Demonstration



Human Attribute Recognition

Facial Attributes Recognition

- Gender
- Age
- Hair Color
- Hair Length
- Hair Style
- Lip Color
- Eyeglasses



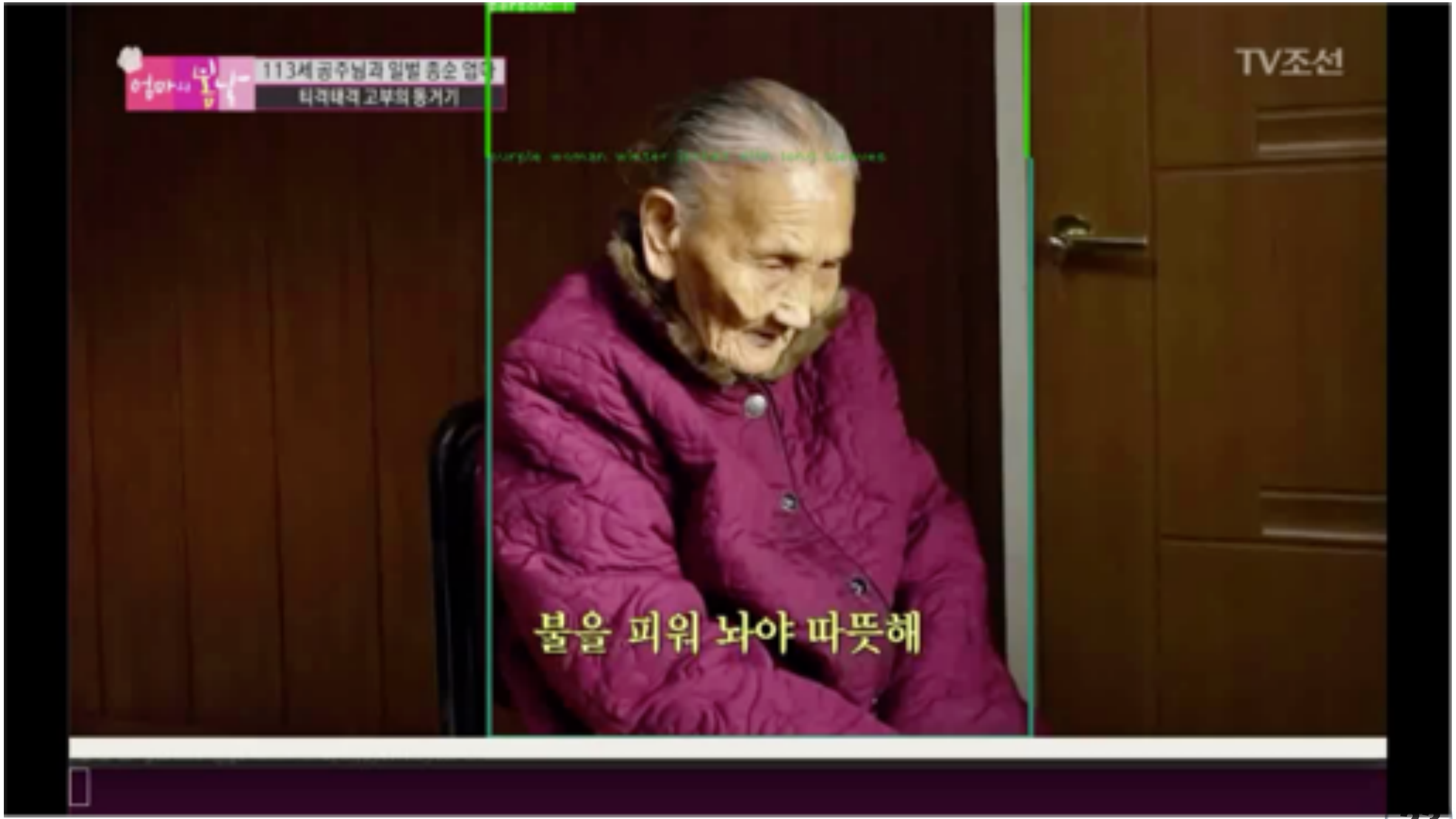
```
Gender : Male
Age : 15.8
Hair Information
 - Color : BLACK(80.16%), BROWN(9.28%)
 - Detail : Strong_Black_Hair_Color
Hair Length : Short
Hair Style : Straight
Lipstick Color : BROWN(100%)
```

psw_d 87.36%

Human Attribute Recognition

Outfit/Accessories Recognition

- Cloth Class
- Sleeve Length
- Cloth Color
- Season
- Accessories

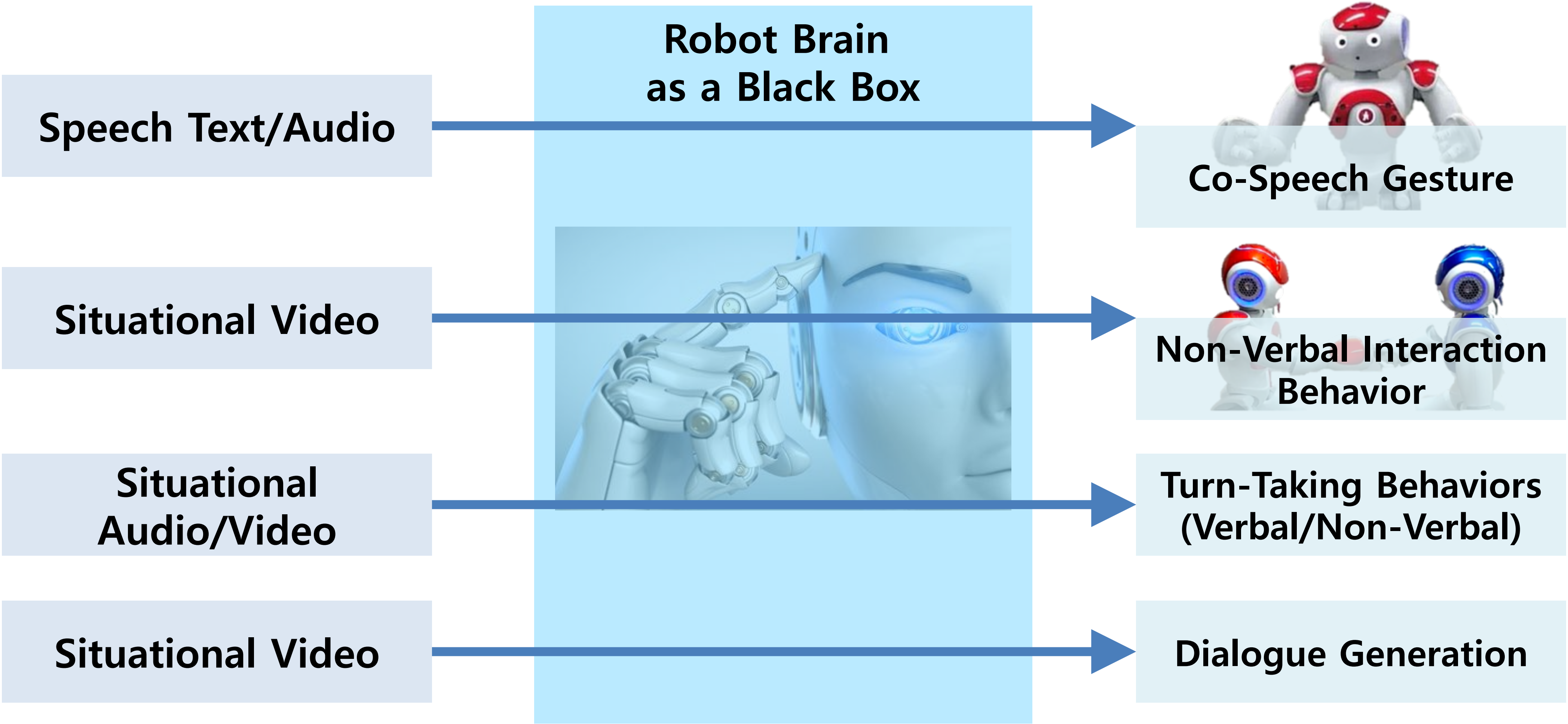


엄마의 봄날
113세 공주님과 일벌 증손 엄마
티격태격 고부의 동거기
TV조선
purple woman winter jacket with long sleeves
불을 피워 놔야 따뜻해

Robot Social AI

Learning-based Approach for Robot Social AI

End-to-End Learning from Human-Human Interaction for Social Situation Awareness and Response Generation



Robot Brain
as a Black Box

Speech Text/Audio → Co-Speech Gesture

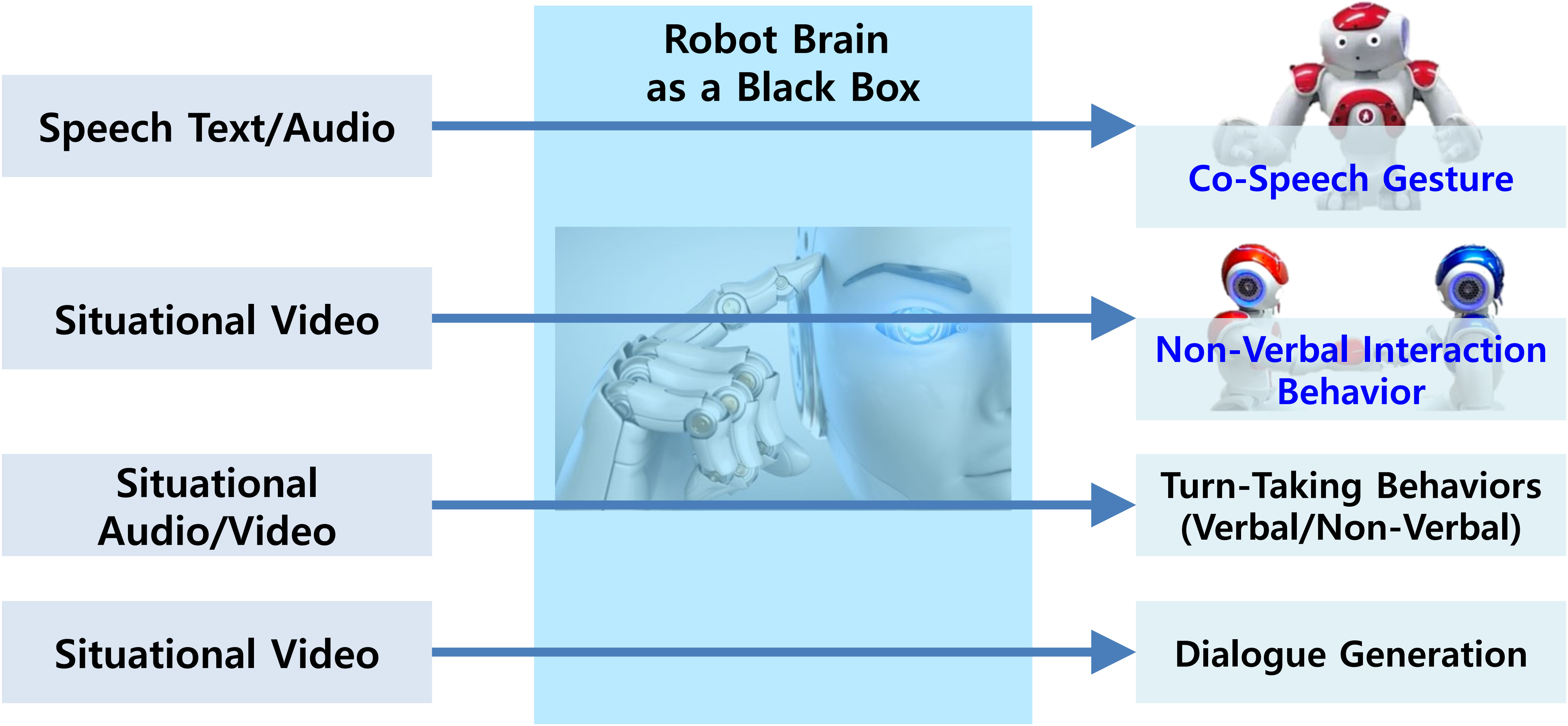
Situational Video → Non-Verbal Interaction Behavior

Situational Audio/Video → Turn-Taking Behaviors (Verbal/Non-Verbal)

Situational Video → Dialogue Generation

Learning-based Approach for Robot Social AI

End-to-End Learning from Human-Human Interaction for Social Situation Awareness and Response Generation



Robot Brain as a Black Box

Speech Text/Audio → Co-Speech Gesture

Situational Video → Non-Verbal Interaction Behavior

Situational Audio/Video → Turn-Taking Behaviors (Verbal/Non-Verbal)

Situational Video → Dialogue Generation

Co-Speech Gesture Generation

I have people skills; I am good at dealing with people.
Can't you understand that? What the hell is wrong with you people?

- One of the key elements of social interaction

  Evaluation of Social Interaction (ESI) Assessment[1]

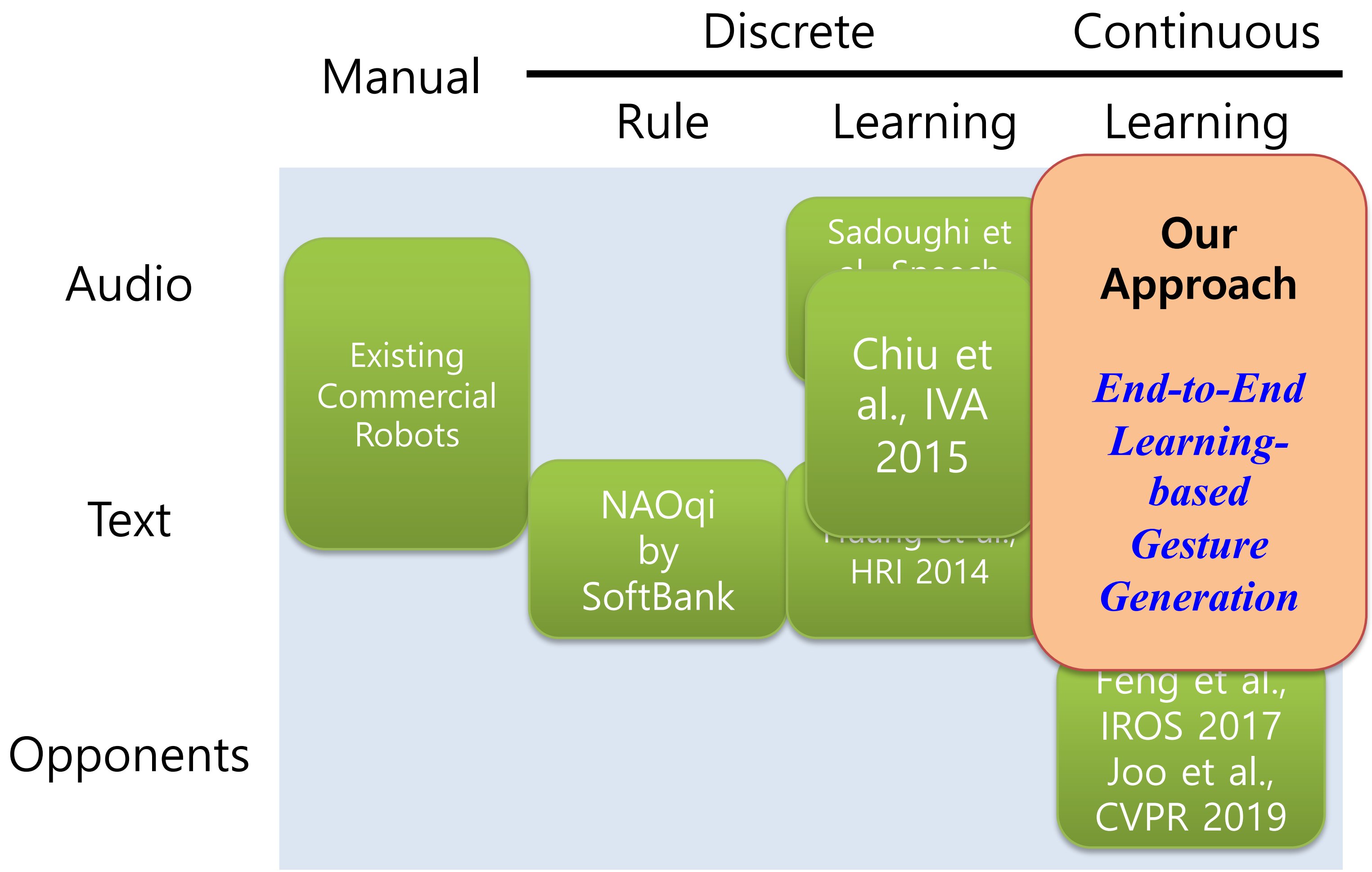
  – Approaches, Gaze, Conversation flow, **Gesture**, Facial expression, …
- More Attention[2], Help listeners comprehend[3], Human likeness

[1] Fisher, A.G. and Griswold, L.A., 2010. Evaluation of social interaction (ESI). Fort Collins, CO.

[2] Bremner, P., Pipe, A.G., Melhuish, C., Fraser, M. and Subramanian, S., 2011, October. The effects of robot-performed co-verbal gesture on listener behaviour. In *2011 11th IEEE-RAS International Conference on Humanoid Robots*.

[3] Cassell, J., McNeill, D. and McCullough, K.E., 1999. Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. Pragmatics & cognition.

Co-Speech Gesture Generation



Manual

Discrete

Continuous

Rule

Learning

Learning

Audio

Text

Opponents

Existing Commercial Robots

NAOqi by SoftBank

Sadoughi et al., Speech

Chiu et al., IVA 2015

Huang et al., HRI 2014

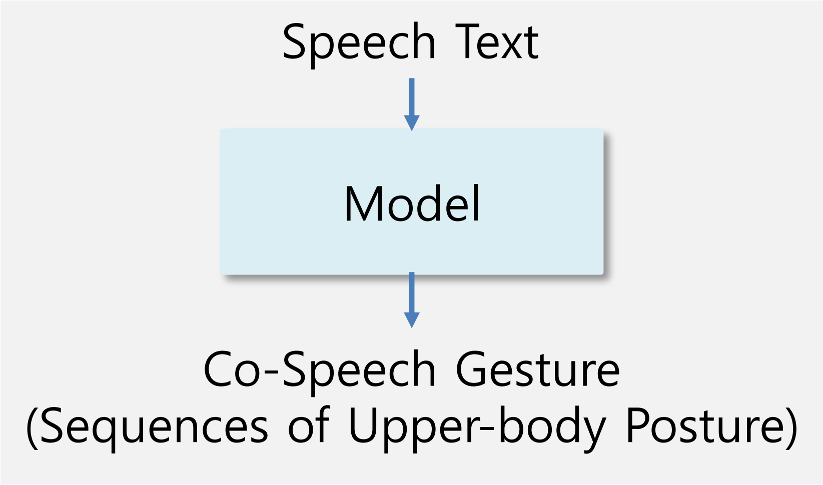
Our Approach

End-to-End Learning-based Gesture Generation

Feng et al., IROS 2017
Joo et al., CVPR 2019

Goal

Generating natural and plausible co-speech gestures for multimodal speech context by end-to-end learning from in-the-wild videos



Speech Text

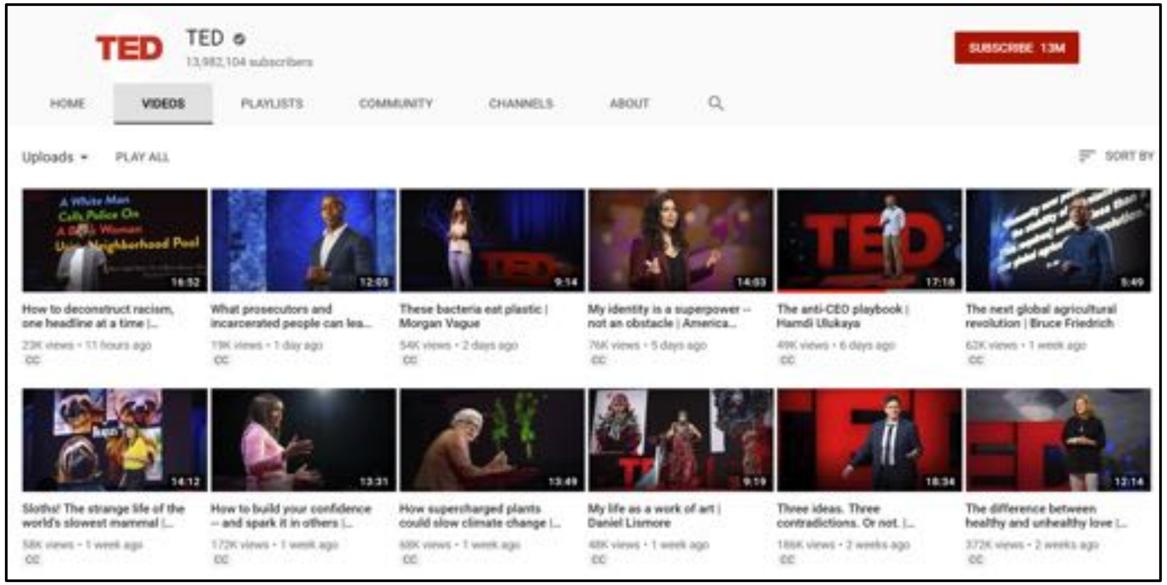
Model

Co-Speech Gesture
(Sequences of Upper-body Posture)

Co-Speech Gesture Generation

Data Acquisition: TED Videos...

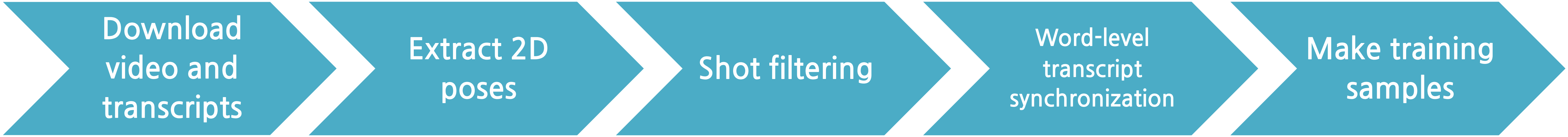
- First **<u>large-scale</u>** & **<u>in-the-wild</u>** dataset
- Why TED talks?
  - – Large enough
  - – Various speech content and speakers
  - – Expect that the speakers use proper hand gestures
  - – Favorable for automation of data collection and annotation


TED
TED
13,982,104 subscribers
SUBSCRIBE 13M
HOME
VIDEOS
PLAYLISTS
COMMUNITY
CHANNELS
ABOUT
Uploads
PLAY ALL
SORT BY
A White Man Calls Police On A Black Woman Using Neighborhood Pool
16:52
How to deconstruct racism, one headline at a time |...
23K views • 11 hours ago
CC
12:05
What prosecutors and incarcerated people can lea...
19K views • 1 day ago
CC
9:14
These bacteria eat plastic | Morgan Vague
54K views • 2 days ago
CC
14:03
My identity is a superpower -- not an obstacle | America...
76K views • 5 days ago
CC
17:18
The anti-CEO playbook | Hamdi Ulukaya
49K views • 6 days ago
CC
5:49
The next global agricultural revolution | Bruce Friedrich
62K views • 1 week ago
CC
14:12
Sloths! The strange life of the world's slowest mammal |...
58K views • 1 week ago
CC
13:31
How to build your confidence -- and spark it in others |...
172K views • 1 week ago
CC
13:49
How supercharged plants could slow climate change |...
68K views • 1 week ago
CC
9:19
My life as a work of art | Daniel Lismore
48K views • 1 week ago
CC
18:34
Three ideas. Three contradictions. Or not. |...
186K views • 2 weeks ago
CC
12:14
The difference between healthy and unhealthy love |...
372K views • 2 weeks ago
CC

Co-Speech Gesture Generation

Data Acquisition: Automated Data Collection Pipeline

Automated Process


Download video and transcripts

Extract 2D poses

Shot filtering

Word-level transcript synchronization

Make training samples

Excluded samples

Co-Speech Gesture Generation

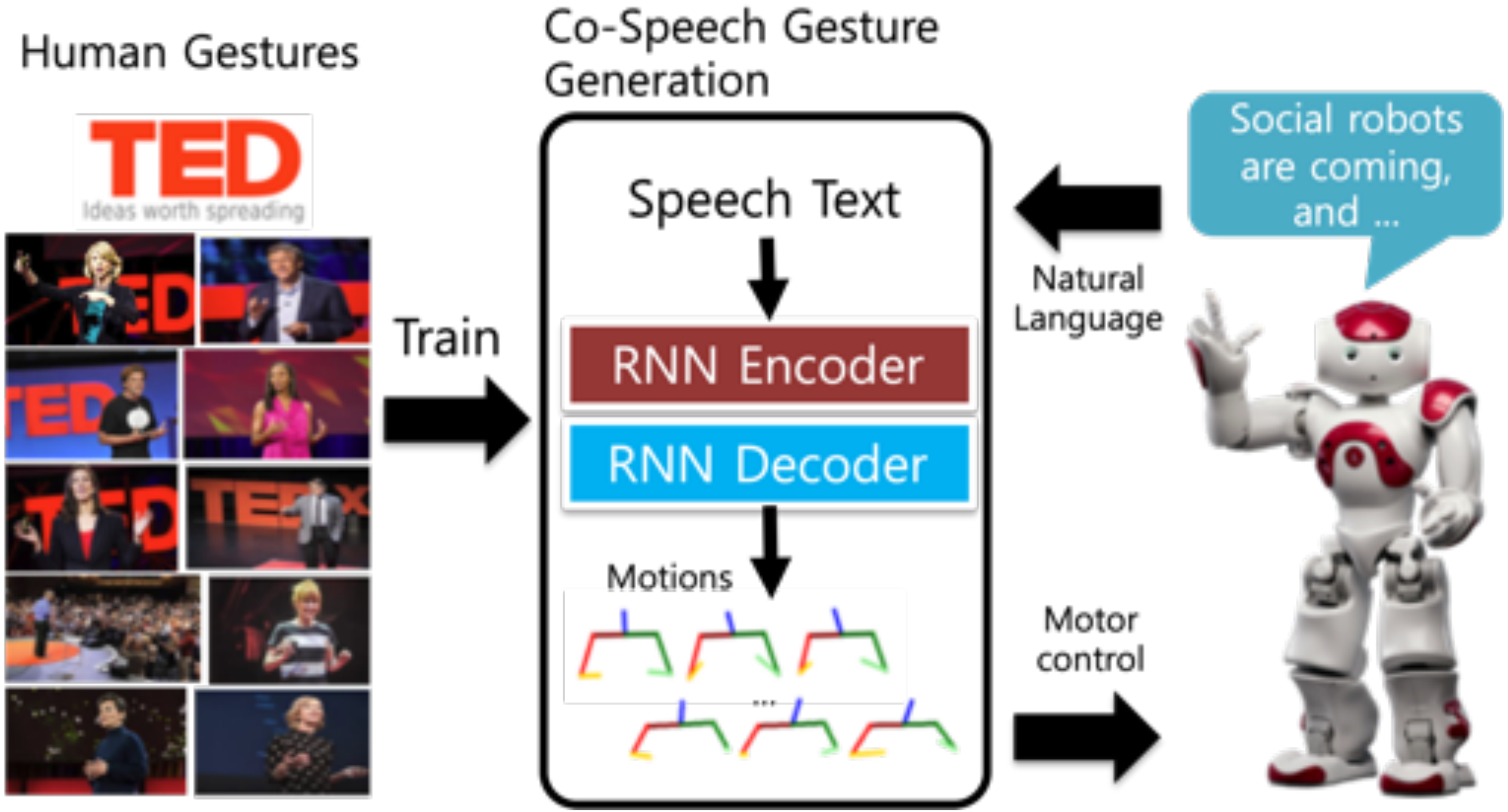
Data Acquisition: Youtube TED Gesture dataset

Number of videos	1,766
Average length of videos	12.7 min
Shots of interest	35,685 (20.2 per video on avg.)
Ratio of shots of interest	25% (35,685 / 144,302)
Total length of shots of interest	106.1 h

Publicly available http://ai4robot.github.io/datasets

Co-Speech Gesture Generation

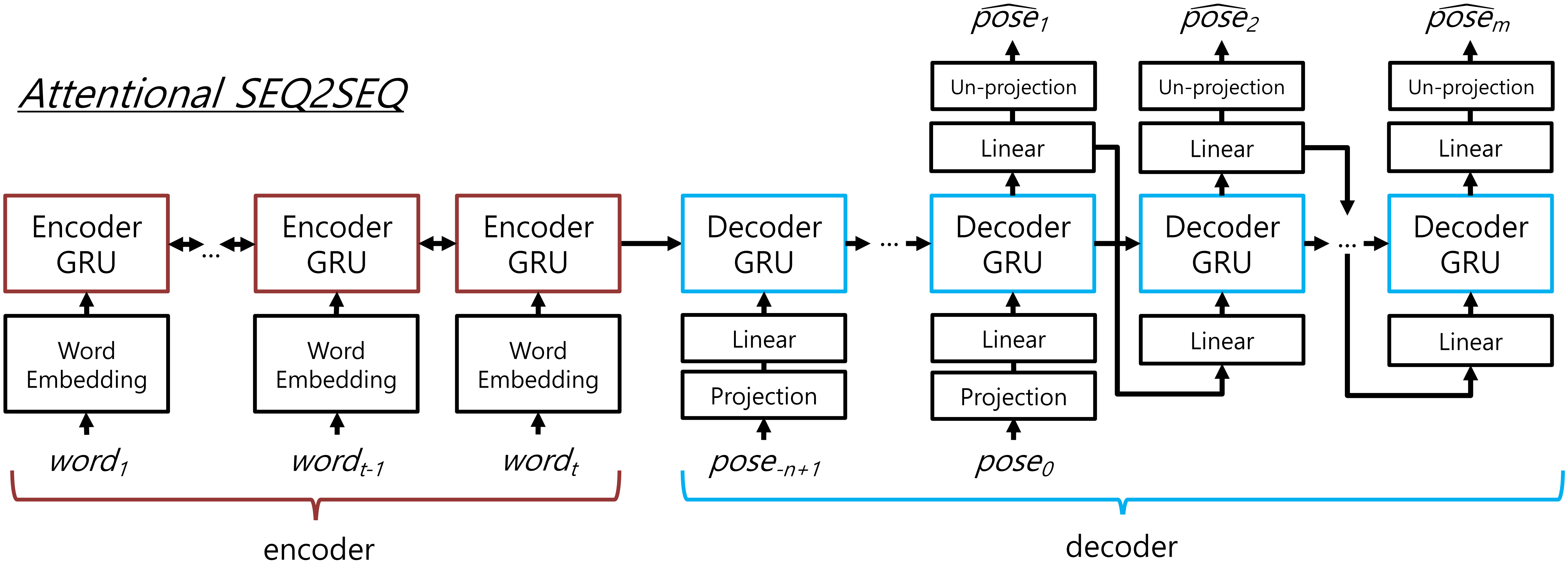
System Architecture



Yoon, Y. et al., Robots Learn Social Skills: End-to-End Learning of Co-Speech Gesture Generation for Humanoid Robots, in the Proc. of The International Conference in Robotics and Automation (ICRA 2019).

Co-Speech Gesture Generation

Deep Text-to-Gesture Generation Model



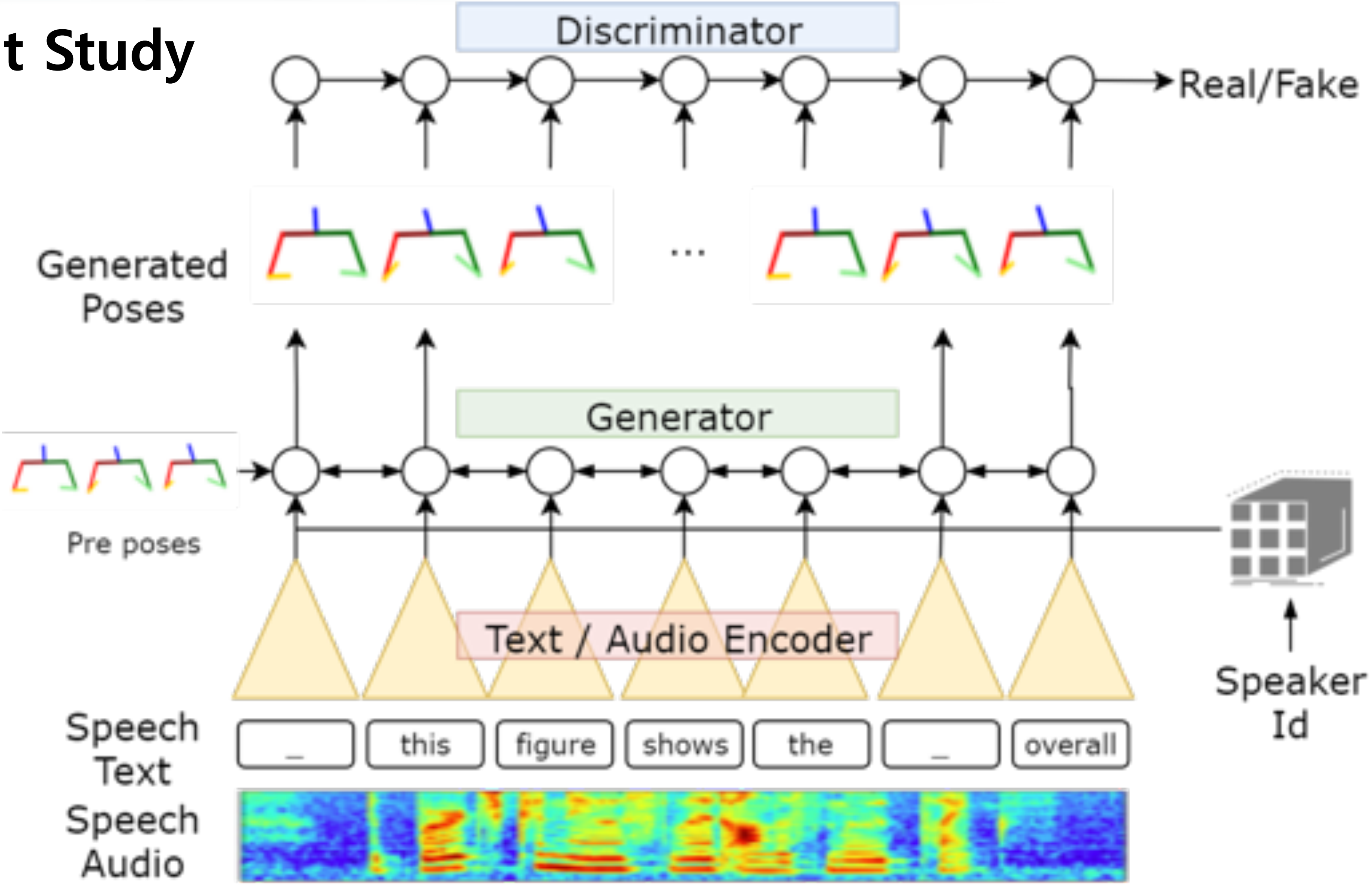
Co-Speech Gesture Generation

Robots Learn Social Skills: End-to-end Learning of Co-Speech Gesture Generation for Humanoid Robots

Youngwoo Yoon, Woo-Ri Ko, Minsu Jang, Jaeyeon Lee, Jaehong Kim, and Geehyuk Lee

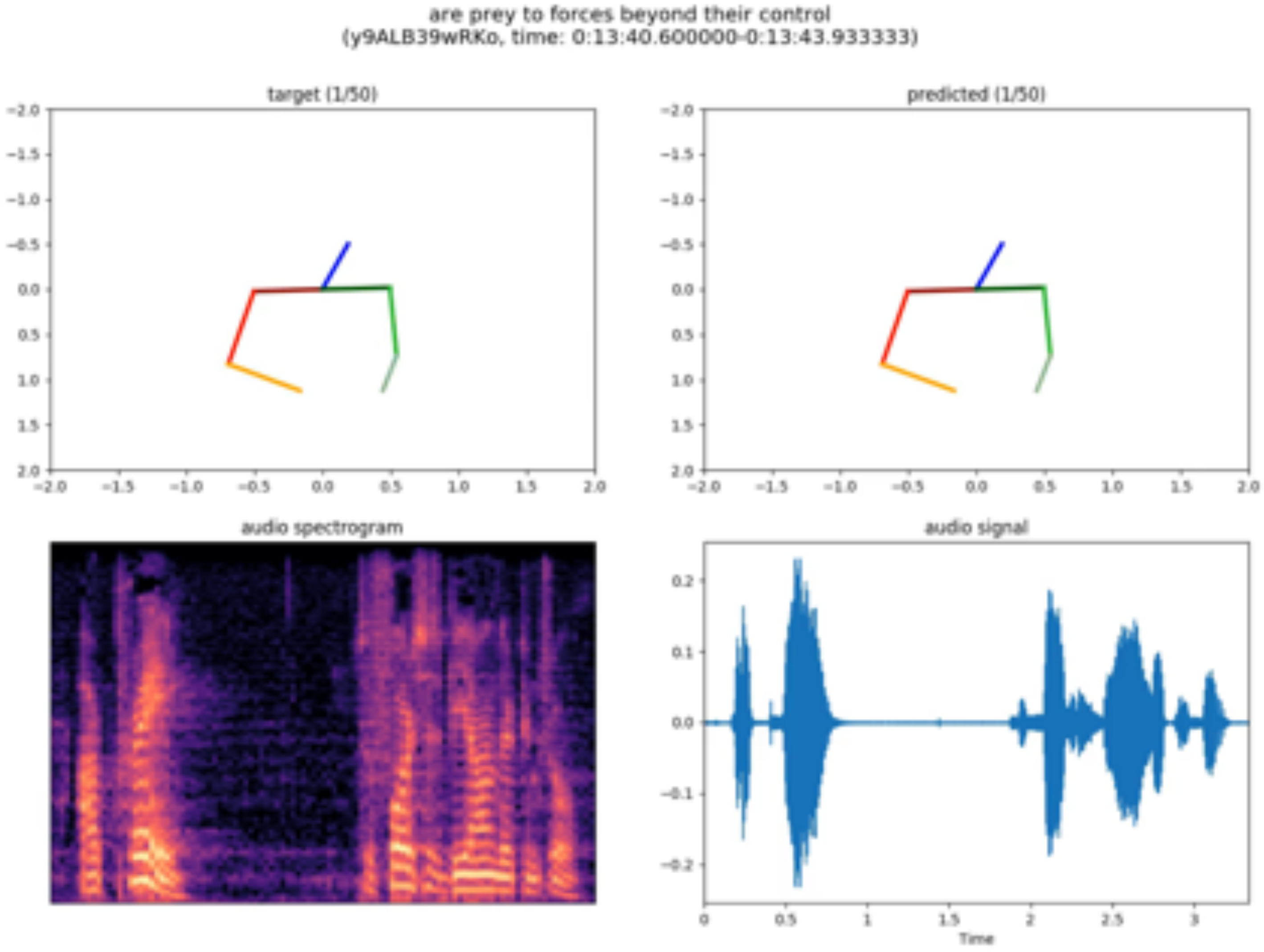
Co-Speech Gesture Generation

Recent Study



Co-Speech Gesture Generation

Recent Study



are prey to forces beyond their control
(y9ALB39wRKo, time: 0:13:40.600000-0:13:43.933333)

target (1/50)

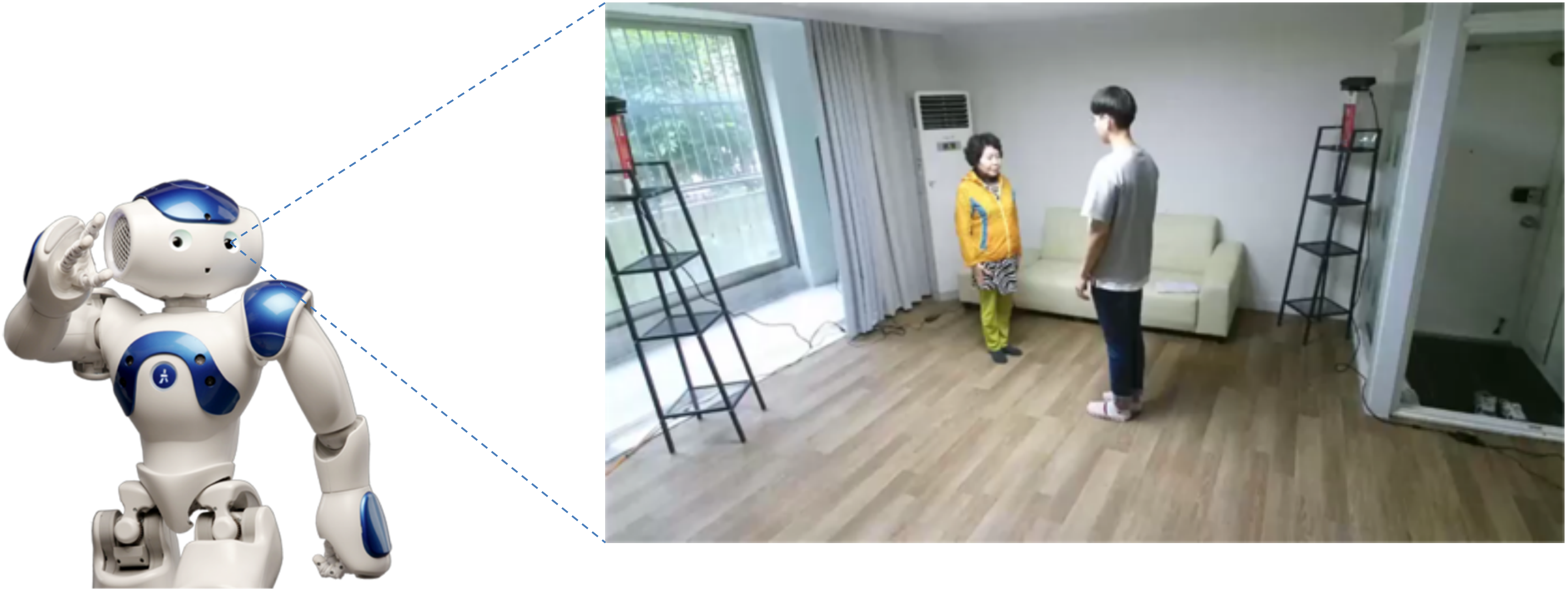
predicted (1/50)

audio spectrogram

audio signal

Act2Act: Non-Verbal Interaction Behavior Generation

Learning to decide
when and how to perform which interaction behavior
by observing human-human interactions



Act2Act: Non-Verbal Interaction Behavior Generation

Data Acquisition: Human-Human Interaction at the testbed

- Participants: 100 elderly people (age > 65)
- Data Format: RGB/Depth/Skeleton/Robot Joint Angles
- Data Scale: 7,500 sets of data
  - 100 interaction groups x 10 scenarios x 5 repetitions x 3 views
  - 500GB

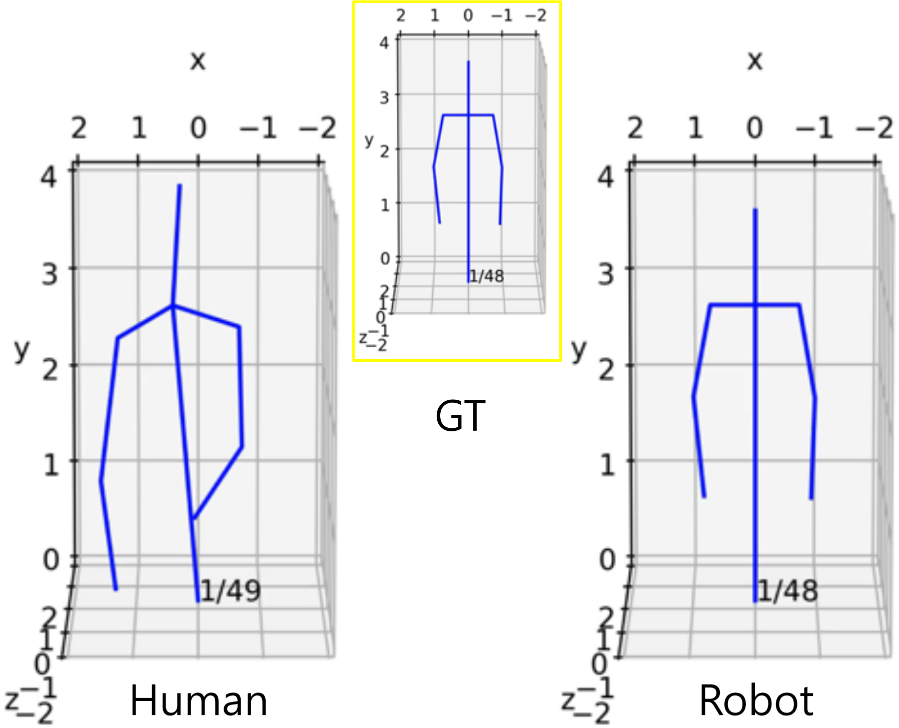
Partially publicly available http://ai4robot.github.io

Act2Act: Non-Verbal Interaction Behavior Generation

Learning Model

Act2Act: Non-Verbal Interaction Behavior Generation

Intermediary Results

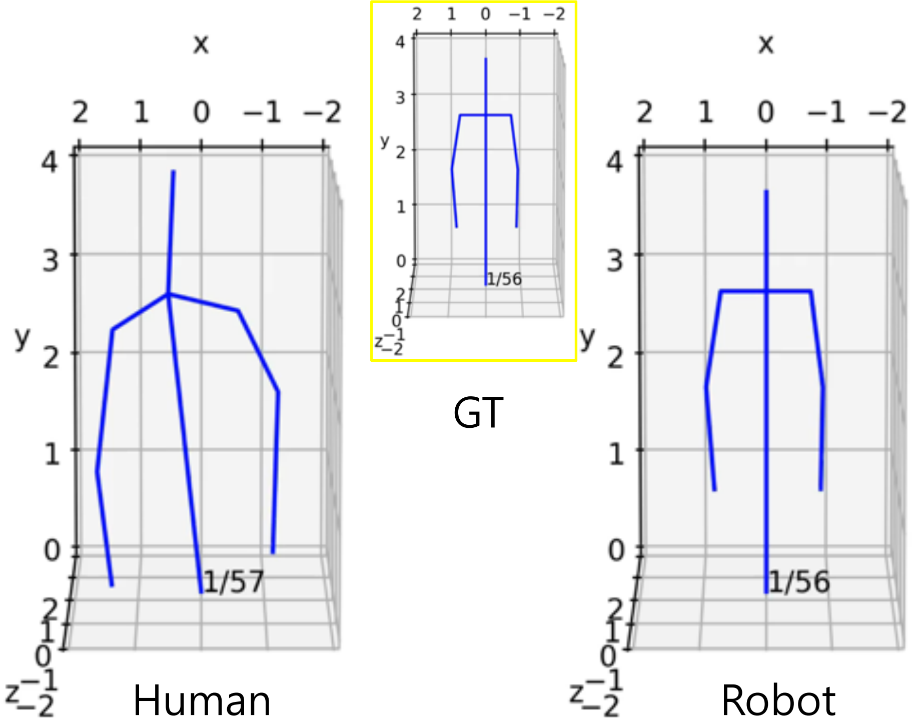


Human

GT

Robot

C001P001A005S004_gen.mp4



Human

GT

Robot

C001P001A006S005_gen.mp4

Summary: Datasets…

Datasets	Org.
Co-Speech Gesture Generation: 1,766 TED video clips, 106.1 hrs of RGB video clips & skeleton data	ETRI
Elderly's Daily Activity Detection: 100 participants (50 elderly, 50 young adults), 112,665 RGBDS video clips	ETRI
Object Instance Registration/Detection: 15 object classes, 830 RGBD video clips	ETRI
Act2Act: Non-Verbal Interaction Behavior Generation: 100 elderly participants, 15,000 RGBDS video clips	ETRI
Turn-Taking Intention Detection: 100 elderly participants, 33 hrs of annotated RGB video clips	ETRI
Long-term Daily Activity: 8 Living Labs, 168,890 motion/wearable/IoT sensor recordings	KETI
ADL Reasoning: 3 Living Labs, 660 hrs of percept sequences and ADL intention annotations	SSU
Elderly Voice: 400hrs of elderly's dialog voice data	MINDsLab

Summary: We're in the 3rd year out of 5 year duration

- Please watch out for open-source software and public datasets in the domain of social robotics and elderly care.

http://ai4robot.github.io

Thank you!