

MULTIMODAL ASSISTED LIVING ENVIRONMENT

LALE AKARUN¹, ALEXEY KARPOV², HULYA YALCIN¹, ALEXANDER
RONZHIN², İSMAIL ARI¹, BARIŞ EVRİM DEMİRÖZ¹, AYSUN ÇOBAN³

¹Computer Engineering, Bogazici University

²Institution of the Russian Academy of Sciences St. Petersburg Institute for Informatics and
Automation of RAS, SPIIRAS

³Electrical & Electronics Engineering, Bogazici University

Audio Signal
Processing

Leaders

Video Signal
Processing



Alexander Ronzhin



Lale Akarun



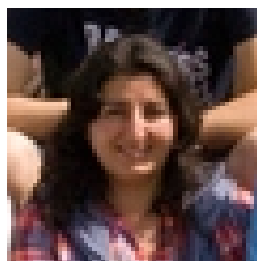
İsmail Arı



Alexey Karpov



Barış Evrim
Demiröz



Hulya Yalcin



Aysun
Coban

Work Plan

| | Week 1 | Week 2 | Week 3 | Week 4 |
|--|-------------------|-------------------|------------------|----------------|
| WP1. Design of the overall system | Completed on time | Completed on time | | |
| WP2. Video detection and tracking system | Completed on time | Completed on time | Almost completed | |
| WP3. Speech recognition | Completed on time | Completed on time | | |
| WP4. System Integration and Module testing | | | Upcoming tasks | Upcoming tasks |
| Documentation | | Almost completed | | Upcoming tasks |

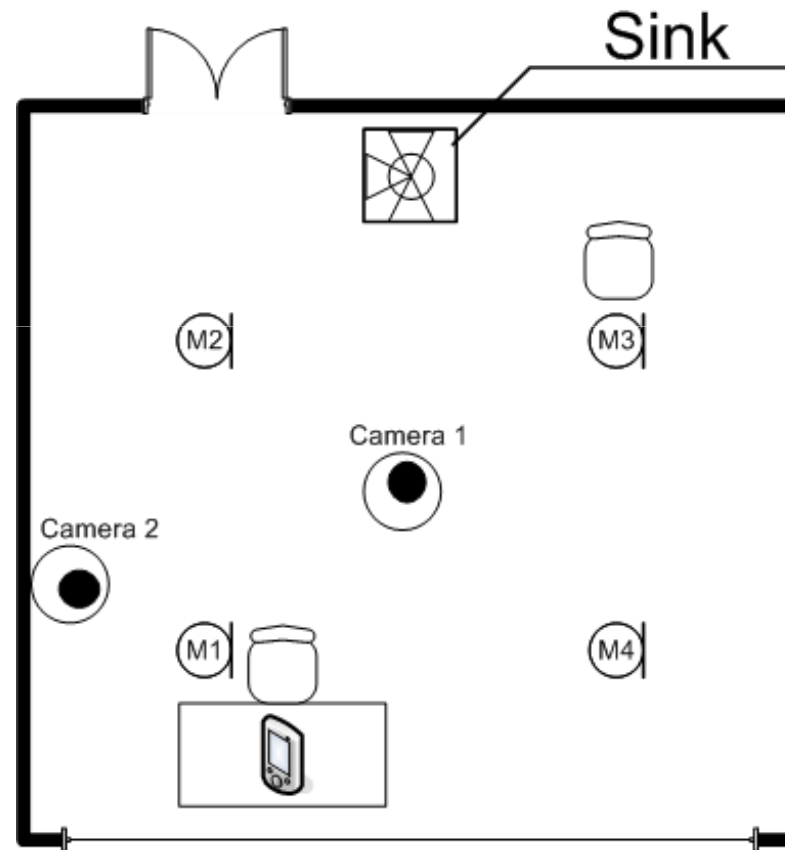
Completed on time
Almost completed
Upcoming tasks

Objectives



- Establish multimodal smart environment
- For elderly people

Sensor Configuration



Omni-directional Cameras



Omni-directional Cameras

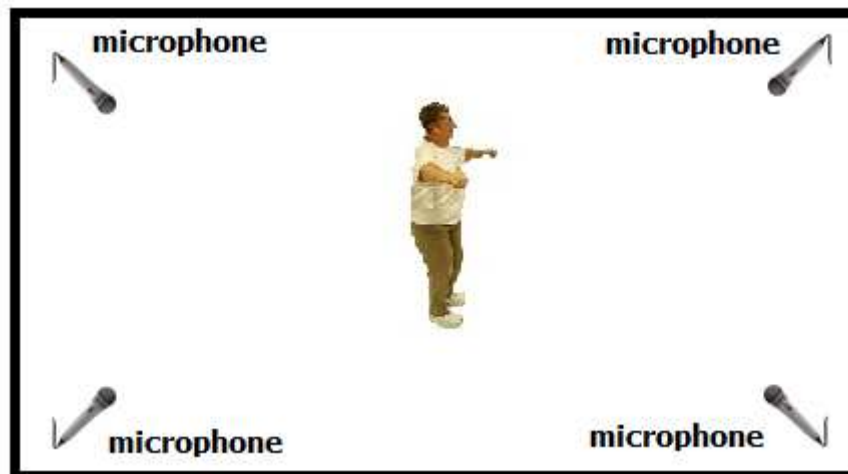


- Wide field of view
- No moving parts
- No blind spots

Microphone Array

Speech is the most natural communication channel for human-computer interaction. Speech synthesis is effective for communicating information and messages to a user without hearing disabilities. Speech recognition is especially useful for people who have difficulty using their hands that preclude using conventional computer input devices.

Speech recognition performance is degraded by the ambient noise and other sound sources in the application environment. Microphone arrays can be used to alleviate this affect by speaker sound localization and directional sound signal reception using beamforming technique.



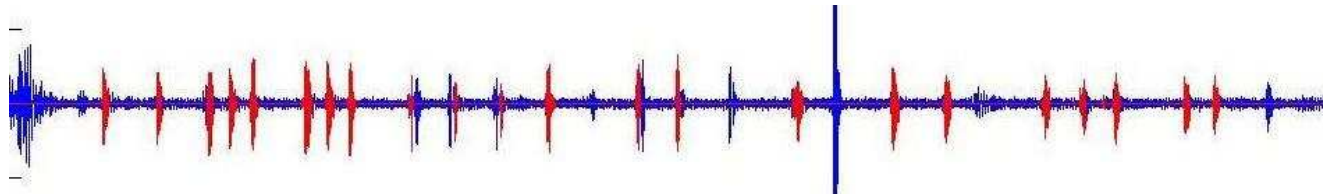
Sound Data Base

- 🔊 Answer Phone
- 🔊 Cough
- 🔊 Moan (Cry after Fall)
- 🔊 Fall

etc.

Speech Recognition

- Voice Activity Detection(VAD)
 - ▣ Applied as a preprocessing step for continuous speech recognition to eliminate false triggers
 - ▣ Three VAD systems have been implemented
 - Rabiner's Method: Energy level and zero-crossing rates of the acoustic waveform
 - Unsupervised learning: Bi-gaussian modeling of the energy level of the signal
 - Supervised learning: Finding the optimal threshold between single Gaussian distributions or GMMs



Speech Recognition Details



- HTK toolkit – HMM based speech recognition

Speech Recognition Details



- Sound database
 - 19 Events
 - 5 English: Yes, No, Answer Phone, Help, Problem
 - 5 Russian: Da, Net, Otvetit, Pomogite, Problema
 - 9 acoustic: Cough, Cry (Moan), Door (open/close), Fall, PhoneRing, Step, Water
 - +1 noise class
 - 100 instances per event

Evaluation of audio processing



- Water class is problematic
 - ▣ Due to low energy signal
- More training data needed for some classes

Human activity detection

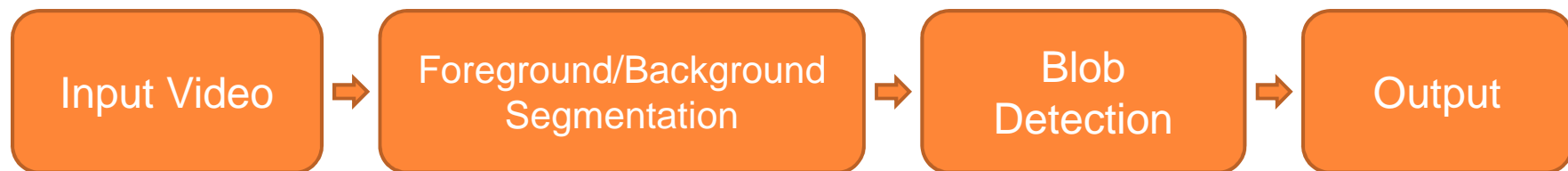
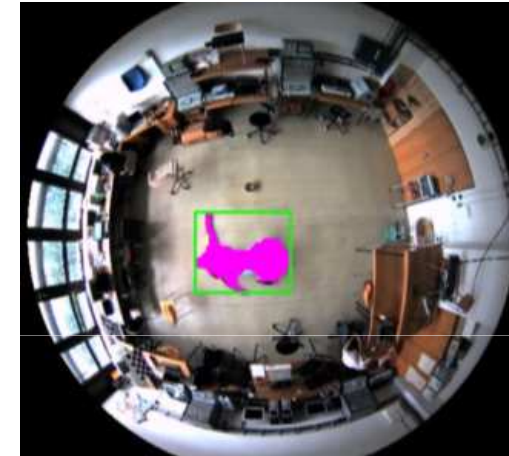


- Human detection
 - ▣ Motion based background modeling (subtraction)

- Human Tracking
 - ▣ Interest point detection and description
 - FAST

- Side effect: Mini software for region labeling on video

Bg/Fg Segmentation and Blob Tracking



Mixture of Gaussians

Bounding Box on Largest Connected Component

Scenario



Enter room
from door
(open &
close)



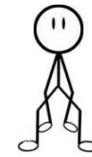
Walk to
table 1



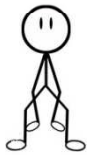
Pick up glass
of water from
table 1



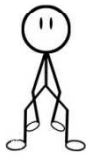
Walk to
chair 1



Sit on
chair 1



Drink
water



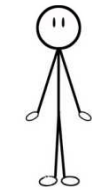
Cough after
drinking



Stand
up



Walk to
table 1



Release
glass



Walk to
sink



Wash
hands



Exit room
(open & close)

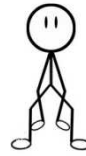
Scenario cont'd



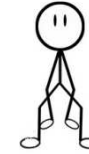
Enter room
from door
(open &
close)



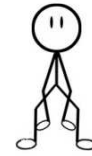
Walk to
chair 2



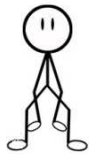
Sit on
chair 2



Phone
rings on
table 2



Say
"Answer
phone"



Talk on
the phone



Stand
up



Walk to
table 1



Pick up
metallic cup
from table 1



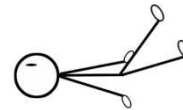
Free
walk



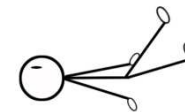
Fall the cup on
the floor and
leave it there



Free walk



Fall



Cry for "help"





Upcoming



- Fisheye camera calibration
- Inter-camera association of points
- Multiple people scenario

Questions?

