# HCI RESEARCH AT BAU: Affect Recognition, Human Behavior Analysis, Mobile Localization and Robotics


Prof. Cigdem Eroglu Erdem

Department of Electrical and Electronics Engineering

Bahcesehir University

İstanbul, Turkey

cigdem.eroglu@eng.bahcesehir.edu.tr

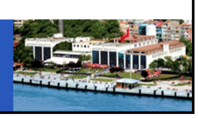
www.bahcesehir.edu.tr



TGMIS 2014, Turkish-German Multimodal Interaction Summit
İstanbul,Turkey, 11 Novemver 2014, http://www.tgmis.itu.edu.tr/


## OUTLINE

- PART I: Affect Recognition
  - Dept. Electrical and Electronics Engineering (Cigdem Eroglu Erdem)
- PART II: Human Behavior Understanding
  - Dept. Software Engineering (Nafiz Arica)
- PART III: Mobile Localization
  - Dept. Computer Engineering (Egemen Ozden)
- PART IV: Human Centered Robotics
  - Dept. Mechatronics Engineering (Berke Gur)



# PART I: Affect Recognition

- Human-Human Interaction
  - Verbal messages
  - Non-verbal messages
    - Reinforce or modify what is said in words
    - Convey information about emotional/mental state
      - Facial expressions
      - Changes in our voice
      - Other bodily signals (Body gestures, heart rate, skin conductance)

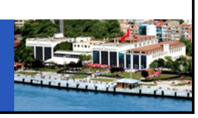
BAHÇEŞEHİR ÜNİVERSİTESİ

# PART I: Affect Recognition

- Human-robot interaction scenarios will involve affect recognition and sythesis.
- Goal: Spontaneous Affect Recognition from Facial Expressions and Speech*
  - Collection of naturalistic audio-visual databases
    - Induced in laboratory (BAUM-1 database)
    - From movies (BAUM-2 database)
  - Facial expression recognition
  - Audio-visual affect recognition

BAHÇEŞEHİR ÜNİVERSİTESİ

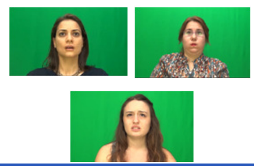
* Project was funded by Technical and Scientific Research Council of Turkey (TUBITAK-1001)



## PART I: Affect Recognition

# BAUM-1 Database*

- Six basic emotions (happiness, sadness, anger, fear, surprise, disgust)
- Other emotions (boredom, contempt)
- Mental states
  - Concentrating
  - Thinking
  - Confused /unsure
  - Interest/curiosity
- Elicited in laboratory using a stimuli video
- In Turkish, 1184 clips
- Recordings are frontal stereo, mono half-profile.
- Baseline FER results: 30%



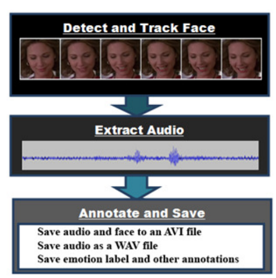

BAHÇEŞEHİR ÜNİVERSİTESİ

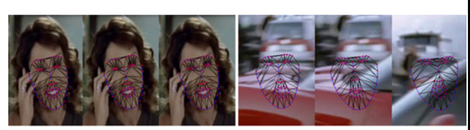
* Web site: baum1.bahcesehir.edu.tr

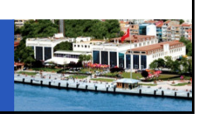
## PART I: Affect Recognition

# BAUM-2 Database*

- Extract emotional facial clips from movies.
  - Detect & track face until a scene cut
  - Improved face tracker
- Multilingual, 1047 clips, six basic emotions
- More naturalistic as compared to acted databases.
- Image based database: BAUM-2i
- Baseline FER results
  - Audio is noisy
  - 57% on BAUM-2i, 49% BAUM-2 (video)





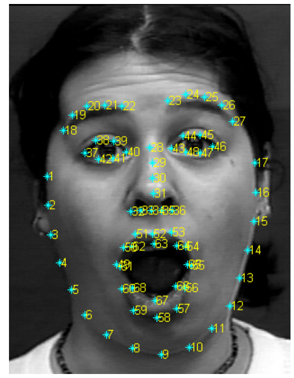
BAHÇEŞEHİR ÜNİVERSİTESİ

* Erdem, Turan, Aydin, «BAUM-2: A Multilingual Audio-Visual Affective Face Database», Multimedia Tools and Applications, 2014. Web site: baum2.bahcesehir.edu.tr

PART I: Affect Recognition

## Facial Expression Recognition by Estimation of the Neutral Face Shape*

- Goal:
  - Alleviate the identity related information in an expressive face image.
  - Increase the facial expression recognition rate.
  - How can we estimate the ID related info (i.e. the neutral face shape)?

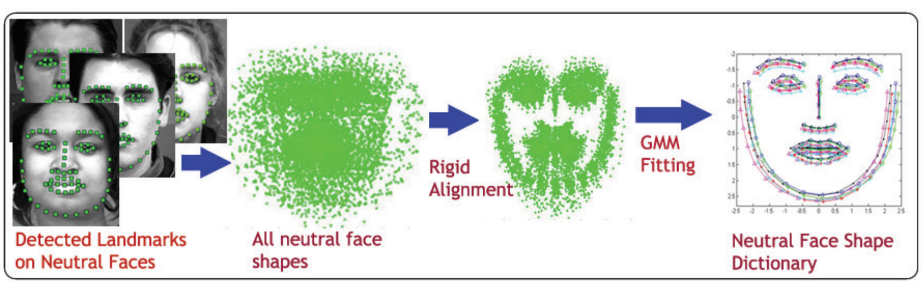


* Ulukaya and Erdem «Gaussian Mixture Model Based Estimation of the Neutral Face Shape for Emotion Recognition», Digital Signal Processing, 2014.

PART I: Affect Recognition

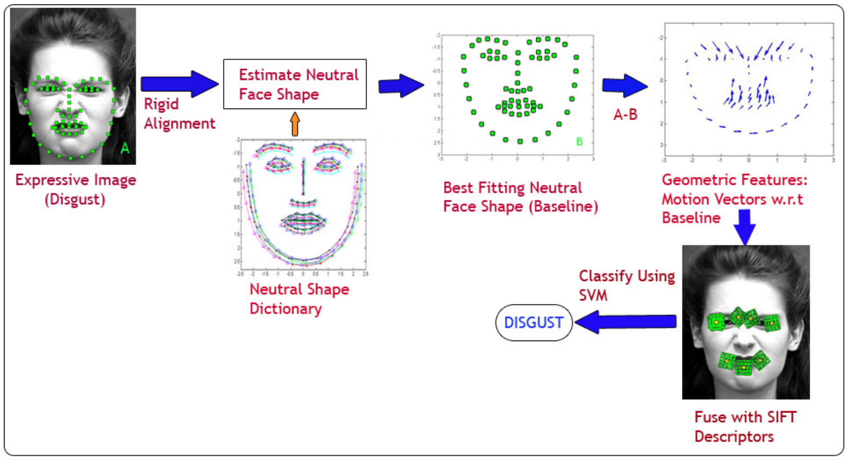
## Facial Expression Recognition by Estimation of the Neutral Face Shape*

- Train a dictionary of neutral face shapes



PART I: Affect Recognition

## Facial Expression Recognition by Estimation of the Neutral Face Shape




PART I: Affect Recognition

## Facial Expression Recognition by Estimation of the Neutral Face Shape

- Results

Average FER recognition rates on CK+ database.

| Geometric Features | Accuracy | Geom. + Appear. Features | Accuracy |
|---|---|---|---|
| CBF | 83.43% | CBF + SIFT | 87.40% |
| **CBF-ENS** | **87.82%** | **CBF-ENS + SIFT** | **90.36%** |
| CBF-NS | 93.88% | CBF-NS + SIFT | 95.37% |

Average FER recognition rates on MMI dataset.

| Geometric Features | Accuracy | Geom. + Appear. Features | Accuracy |
|---|---|---|---|
| CBF | 53.34% | CBF + SIFT | 58.00% |
| **CBF-ENS** | **58.93%** | **CBF-ENS + SIFT** | **62.52%** |
| CBF-NS | 66.19% | CBF-NS + SIFT | 67.06% |

PART I: Affect Recognition

## Facial Expression Recognition by Estimation of the Neutral Face Shape

- Results
  - Significant improvement in cross-database experiments

Cross-database validation of the proposed method. The average emotion recognition rates on the MMI dataset are given using the neutral shape dictionary and SVM classifiers trained on the CK+ dataset.

| Features Used | Ave. Emotion Recognition Rate |
|---|---|
| CBF | 31.39% |
| **CBF-ENS** | **53.95%** |
| CBF-NS | 61.72% |

BAHÇEŞEHİR ÜNİVERSİTESİ

PART I: Affect Recognition

## Audio-Visual Affect Recognition Based on Apex Frame Selection*

- Goal: Given an expressive video use the apex frames for recognition.
- Apex frame: frames at which intensity of the facial expression is maximum.

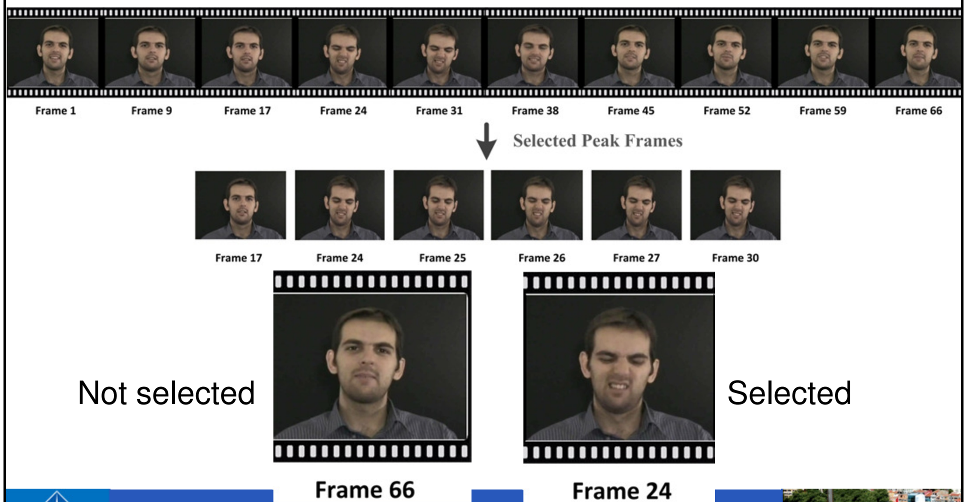
Neutral Onset frame Apex frame (frame16)

- **Problem** **We do not know which frames are the peak frames in a video clip.**

BAHÇEŞEHİR ÜNİVERSİTESİ

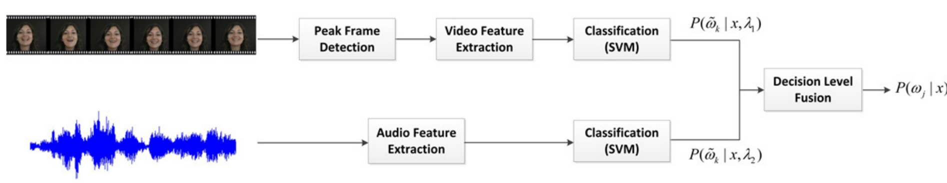
* Zhalehpour, Akhtar, Erdem, «Multimodal Emotion Recognition with Automatic Peak Frame Selection», IEEE Symp. INISTA, June 2014.

## PART I: Affect Recognition
# Peak Frame Selection




BAHÇEŞEHİR ÜNİVERSİTESİ

## PART I: Affect Recognition
# Multimodal Emotion Recognition by Decision Level Fusion




- Image features: LPQ etc.
- Audio features: MFCC + RASTA-PLP
- Audio-visual recognition accuracy on eNTERFACE dataset: 76%

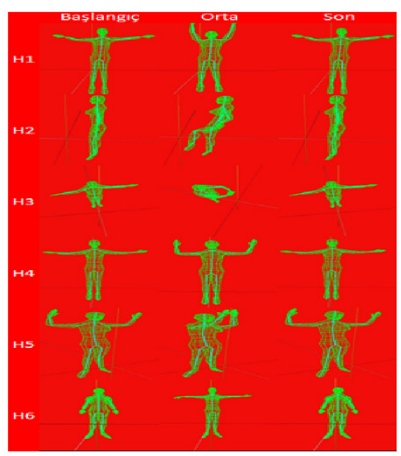
BAHÇEŞEHİR ÜNİVERSİTESİ

# PART II – Human Behaviour Analysis

- Physiotherapy Guidance by Motion Analysis Based on Hidden Markov Model
  - Recep Doğa SİYLİ, Boğaziçi Üniversitesi
  - Lale AKARUN, Boğaziçi Üniversitesi
  - Nafiz ARICA, BAU, nafiz.arica@eng.bahcesehir.edu.tr
- Goal: Physiotherapy guidance at home
- Method
  - Analyze motion data collected by Kinect
  - Compare the performed motion with the pre-stored correct motion model and give feedback to the patient
  - Model motion using various HMMs: left-to-right, circular etc.

# PART II – Human Behaviour Analysis

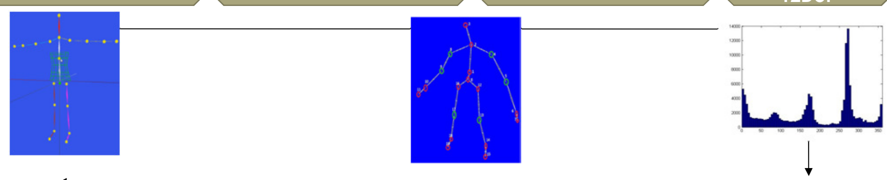
- Data:
  - 6 different gestures
  - 186 gesture sequences
  - 11.418 poses



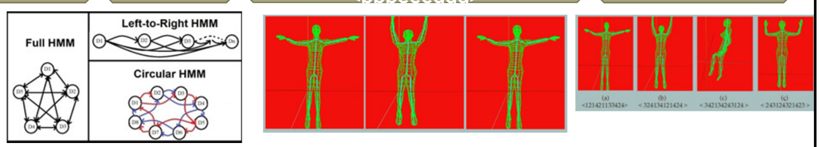

Şekil. 6. Hareketler

# PART II – Human Behaviour Analysis
## Physiotherapy Guidance*

3D Position 20Jointsx3Dim→60DoF → 2D Relative Degrees 20 Jointsx2Dim→40DoF → Feature Selection 6Jointsx2Dim→12DoF → Clustering Degrees 12DoF



GestureRecognition /Verification ← Training HMM ← Forming Motion Sequences of Pose Clusters <bbbcccaaa> ← Clustering of Poses 1DoF <pose->b>

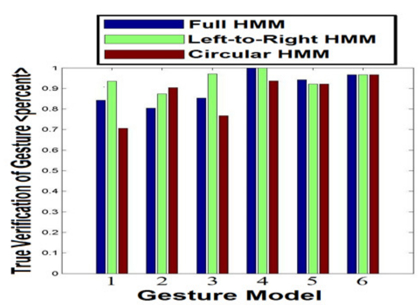



*Siyli, Akarun, Arıca, «Physiotherapy Guidance by Motion Analysis Based on HMM», IEEE Conf. on Signal Processing and Applications (SIU), 2013 .



# PART II – Human Behaviour Analysis
## Physiotherapy Guidance - Results

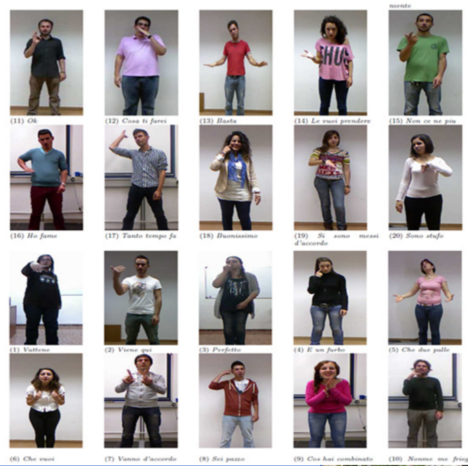
| Gesture Number | Recognition Rate |
|---|---|
| 1 | 93% |
| 2 | 88% |
| 3 | 97% |
| 4 | 99% |
| 5 | 92% |
| 6 | 96% |




- Future Work: give feedback about which part of the gesture has been done incorrectly

# PART II – Human Behaviour Analysis
## Gesture Recognition*

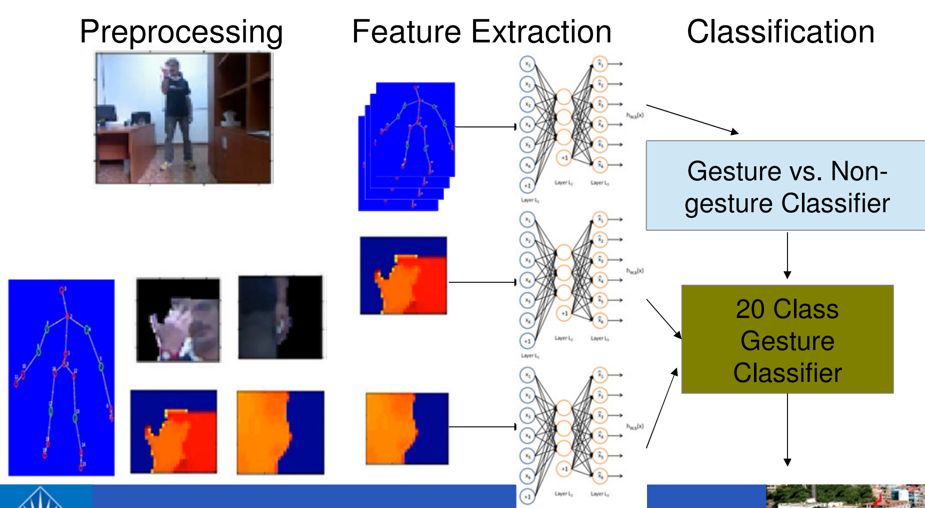
- Aim
  - Gesture spotting in continuous videos
  - Gesture classification (20 Italian gestures)
  - Fusion of multi-modal features from Kinect
    - RGB
    - Depth
    - Skeleton



* ChaLearn 2014 , Challenge and Workshop on Pose Recovery, Action Recognition, Age Estimation and Cultural Event Recognition, http://gesture.chalearn.org/mmdata

BAHÇEŞEHİR ÜNİVERSİTESİ

# PART II – Human Behaviour Analysis
## Gesture Recognition - Method




BAHÇEŞEHİR ÜNİVERSİTESİ

# PART II – Human Behaviour Analysis

## Gesture Recognition - Results

| | Spotted as Gesture (85.4%) | | Classified as Non-Gesture |
|---|---|---|---|
| | **True class** | **Wrong Class** | |
| **Given as gesture** | 74.9% | 12.3 % | 14.6% |
| **Given as non-gesture** | 15% | | |

- Frame-based labeling performance
  - 93% correct labeling for gesture vs. non-gesture classification
  - Winner team acquired 98%
- Future work: Handle with missing data

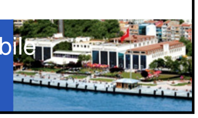
BAHÇEŞEHİR ÜNİVERSİTESİ

# PART III: A Hybrid Framework for Mobile Localization*

- Team:
  - Kemal Egemen Özden (BAU) kemalegemen.ozden@bahcesehir.edu.tr
  - Mehmet Tozlu, Salih Ergüt (Avea Labs)
  - Project funded by Avea and Turkish Ministry of Science and Industry.
- Goal: Combine RF techniques and computer vision methods for accurate localization on mobile phones.
  - GPS: decent accuracy outdoors, fails indoors; GSM: poor localization performance, WiFi: requires dense hotspots
  - Vision: requires offline 3D model generation, matching is computationally intensive on mobile devices

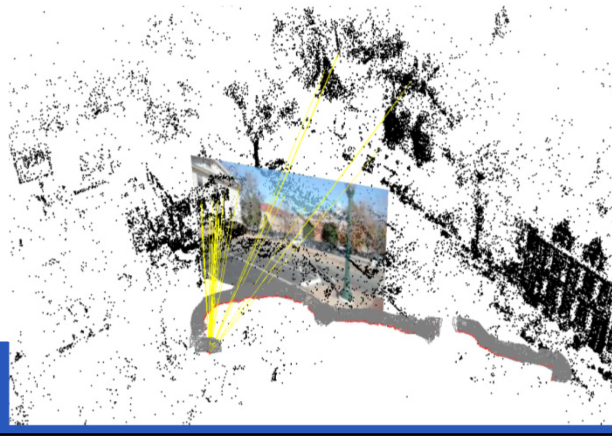
BAHÇEŞEHİR ÜNİVERSİTESİ

*Ozden and Ergut, «A Hybrid Localization Framework for mobile devices», NGMAST, 2014.



# PART III: A Hybrid Framework for Mobile Localization

Intuition: Given a 3D model and a 2D snapshot of it, it is possible to locate the camera position relative to the 3D model (camera external calibration problem)



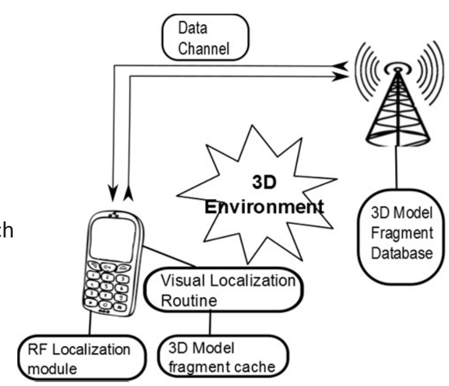
# PART III: A Hybrid Framework for Mobile Localization

- A rough localization from RF methods.
- Use as a query to a remote 3D model fragment server.
- Model fragment is downloaded and cached.
- Image from camera is matched against this small subset of 3D models.

**Scalable**: No need to keep all the model or match the image against all 3D models.

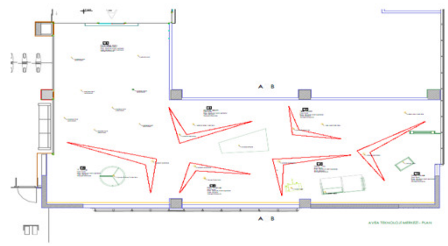
**Accurate**: Vision based results are often more accurate (10 cm to few meters).

<u>This approach has potential for new Augmented Reality and micro navigation applications.</u>



# PART III: A Hybrid Framework for Mobile Localization

- 3D models and floor plans need to be registered as well.

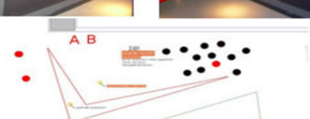


- We developed methods using 3D-2D correspondences (line or point) between 3D model and floorplan.

# PART III: A Hybrid Framework for Mobile Localization

- Augmented Reality Application







# PART IV: Human Centered Robotics Research RoBAUtics Lab

- **Coordinators**:
  - Berke Gür (Mechatronics Engineering) berke.gur@eng.bahcesehir.edu.tr
  - Emel Arican
  - Stanford Artificial Intelligence Laboratory (Prof. Oussama Khatib)
- Realization of highly capable, dexterous, but cost-effective manipulation
- Ability to operate in complex and unstructured environments
- Advanced task and posture based control strategies
- Simultaneous execution of multiple tasks and task prioritization
- Multi-point contact & interaction with the environment
- Learning of & adaptation from human behavior & by experience

STANFORD & BAHÇEŞEHİR

"ROBOTICS RESEARCH PROJECT"

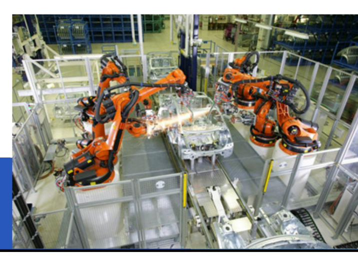
BAU ROBOTICS LAB

www.bahcesehir.edu.tr

BAHÇEŞEHİR ÜNİVERSİTESİ

# PART IV: RoBAUtics Lab Human Friendly Robotics

- Intrinsically safe robots that can co-exist with humans
  - Novel hardware designs
  - Multi-modal perception methods
  - Advanced control strategies
  - Built-in cognition & autonomy
  - Human-robot collaboration

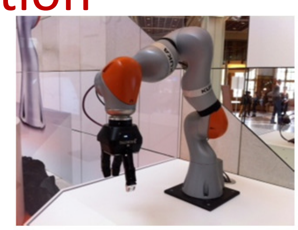
BAHÇEŞEHİR ÜNİVERSİTESİ

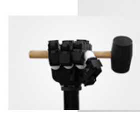
# PART IV: RoBAUtics Lab
## Haptics

- Dynamic & haptic simulation in virtual environments
  - Modeling & rendering stiffness, texture, etc.
- Haptic tele-operation
  - Bandwidth limitations
  - Time delays
- Novel haptic hardware design
- Fusion of haptic feedback with multi-modal sensory perception
  - Vision
  - Aural

BAHÇEŞEHİR ÜNİVERSİTESİ

# PART IV: RoBAUtics Lab
## Object Manipulation





- Dexterous object manipulation
  - Grasping, squeezing, releasing
  - Multi-point contact & multiple constraints
- Operational space, posture & whole body control
- Computer vision
  - Environment mapping
  - Object recognition
- Robot learning
  - Learning by demonstration
  - Modeling & adapting human behavior

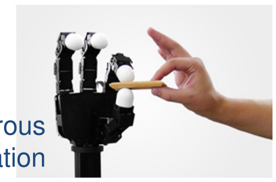
BAHÇEŞEHİR ÜNİVERSİTESİ

# PART IV: RoBAUtics Lab

## Facilities & Research Capabilities

Human Friendly Robotics


Dexterous Manipulation



Operational Space Control

Haptic Interaction

Mobile Manipulation

BAHÇEŞEHİR ÜNİVERSİTESİ

# MAX-PLANCK SCIENCE TUNNEL @İstanbul

www.bilimtuneli.com

Mall of İstanbul

Oct. 18 - Feb. 12

BAHÇEŞEHİR ÜNİVERSİTESİ

- Thank you for your attention…