EE-704

Computational perception using multimodal sensors

Aran Oya, Odobez Jean-Marc

Cursus	Sem.	Type
Electrical Engineering		Obl.

Language of teaching	English
Credits	4
Session	
Exam	Multiple
Workload	120h
Hours	**56**
Courses	32
TP	24
Number of positions	**20**

Frequency

Every 2 years

Remark

Every 2 years. Next time: Fall 2018

Summary

The course will cover perceptual modalities in computers, models for analyzing people (representation, detection an localization, segmentation, tracking, recognition).

Content

1. Perceptual modalities in computers. Vision, hearing, touch, smell. basic fusion principles.
2. Models for analyzing people. introduction to probabilistic graphical models. Basic concepts. Bayesian Networks (BNs). Learning and inference in BNs. Dynamic Bayesian Networks (DBNs). Exact and approximate inference. Examples.
3. Analyzing people. fundamental tasks.
a. Representation. The problem of representation in computational perception. Global vs. local representations. Visual models for faces, heads, hands, and full-bodies (shape/appearance, exemplars, geometric models). Models and features for speech and audio processing.
b. Detection and localization. Basic concepts. Detection as binary classification and as random sampling. Visual localization: skin color modeling, face localization. Audio localization: microphone arrays. Audio-visual fusion for speaker detection.
c. Segmentation. Basic concepts. Visual segmentation: background subtraction. Audio segmentation: source separation, speaker turn segmentation, speaker clustering.
d. Tracking. State space representation. Dynamic modeling. Human motion modeling. Multi-person tracking. Visual, audio and multimodal tracking of people.
e. Recognition. Recognition tasks. Visual recognition: facial expressions, gestures, actions, interaction. Audio recognition: speech, emotion, multi-speaker events. Audio classification. Multimodal recognition: actions.

Keywords

Artificial perception, human representation, multi-modalities, audio, video, probabilistic model, graphical models.

Learning Prerequisites

Recommended courses

Undergraduate-level knowledge of linear algebra, statistics, image and signal processing.

Assessment methods

- written exam
- homeworks (includes practical work)
- paper presentation

Resources

Bibliography

- C. Bishop, Pattern Recognition and Machine Learning, Springer, 2008
- M. Brandstein and D. Ward (eds.), Microphone Arrays, Springer, 2001
- D. Forsyth and J. Ponce, Computer Vision: a Modern Approach, Prentice Hall, 2002
- B. Gold and N. Morgan, Speech and Audio Processing. Wiley, 1999
- M. I. Jordan (ed.), Learning in Graphical Models, MIT Press, 1999

The library recommends:

- "Computer vision : a modern approach / David A. Forsyth, Jean Ponce ; international ed. contrib. by Soumen Mukherjee ... [et al.]". Year:2012. ISBN:978-0-13-608592-8
- "Learning in graphical models / ed. by Michael I. Jordan". Year:1999. ISBN:0-262-60032-3
- "Microphone arrays : signal processing techniques and applications / Michael Brandstein ... [et al.](eds.)". Year:2001. ISBN:978-3-540-41953-2
- "Pattern recognition and machine learning / Christopher M. Bishop". Year:2008. ISBN:978-0387-31073-2
- "Speech and audio signal processing / Ben Gold, Nelson Morgan, Dan Ellis ; with contrib. by Hervé Bourlard ... [et al.]". Year:2011. ISBN:978-0-470-19536-9

Ressources en bibliothèque

- Pattern Recognition and Machine Learning / Bishop
- Microphone Arrays / Brandstein
- Computer Vision: a Modern Approach / Forsyth
- Learning in Graphical Models / Jordan
- Speech and Audio Processing / Gold