EPFL

EE-623   Perception and learning from multimodal sensors

Odobez Jean-Marc

Cursus	Sem.	Type
Electrical Engineering		Opt.

Language of teaching	English
Credits	4
Session	
Exam	Written & Oral
Workload	120h
Hours	**56**
Courses	28
Exercises	10
TP	18
Number of positions	**20**

Frequency

Every 2 years

Remark

Next time: Fall 2022

Summary

The course will cover different aspects of multimodal processing (complementarity vs redundancy; alignment and synchrony; fusion), with an emphasis on the analysis of people, behaviors and interactions from multimodal sensor, using statistical models and deep learning as main modeling tools.

Content

Multimodal processing is at the core of the perception of our world: we see, we hear, we touch, we smell, we taste, and we move. Being able to analyze and combine multimodal streams of data is therefore an inherent ability that should be addressed by artificial intelligence systems, and comprises several core challenges like how to represent multimodal information? How to deal with modality asynchrony? How to fuse complementary or redundant information? How to do Co-training of models?
The course will cover this topic, with a particular emphasis on the analysis of people and of their behaviors, including interactions, from multimodal sensor streams (with a bias towards vision and audio). We will rely on Bayesian statistical models and deep learning as main modeling tools. Within this domain, the course will comprise different lectures, labs, and further reading about the following topics:

1. Introduction to multimodal processing - application and main issues

2. Fundamentals in Bayesian models and deep learning

3. Audio processing: voice analysis, speaker diarization, sound localization

4. Video processing: fundamental tasks - person detection, body, face and hand representation detection and learning; single and multiple person audio-visuel tracking.

5. Multimodal processing: joint representation learning, alignment, fusion

6. Multimodal Activity recognition: gaze and attention models; facial expression and emotion recognition; gesture recognition; social analysis, multi-person and multi-modal action and interaction recognition.

Keywords

Multimodality, human activity analysis, interactions, machine learning, computer vision, audio processing.

Learning Prerequisites

Required courses

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

Recommended courses

Introduction to machine learning course.

Assessment methods

Written and oral.

Resources

Bibliography

Pattern Recognition and Machine Learning, C·. Bishop, Springer, 2008.

Ressources en bibliothèque

• Pattern Recognition and Machine Learning / Bishop