6th International Electronic Conference on Sensors and Applications

HUMAN ACTIVITY RECOGNITION BASED ON DEEP LEARNING TECHNIQUES

Manuel Gil-Martín

Departamento de Ingeniería Electrónica
Universidad Politécnica de Madrid

Contents

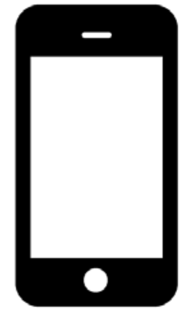
- Introduction
- Materials and Methods
- Results
- Discussion
- Conclusions

Introduction

- Human Activity Recognition (HAR)
  - Recognize different activities performed by a person
- Applications
- On-body sensors vs. smartphones
- Raw data vs. frequency domain
- Deep learning architectures using Convolutional Neural Networks (CNNs)
- This work
  - Comparison of deep learning and signal processing strategies

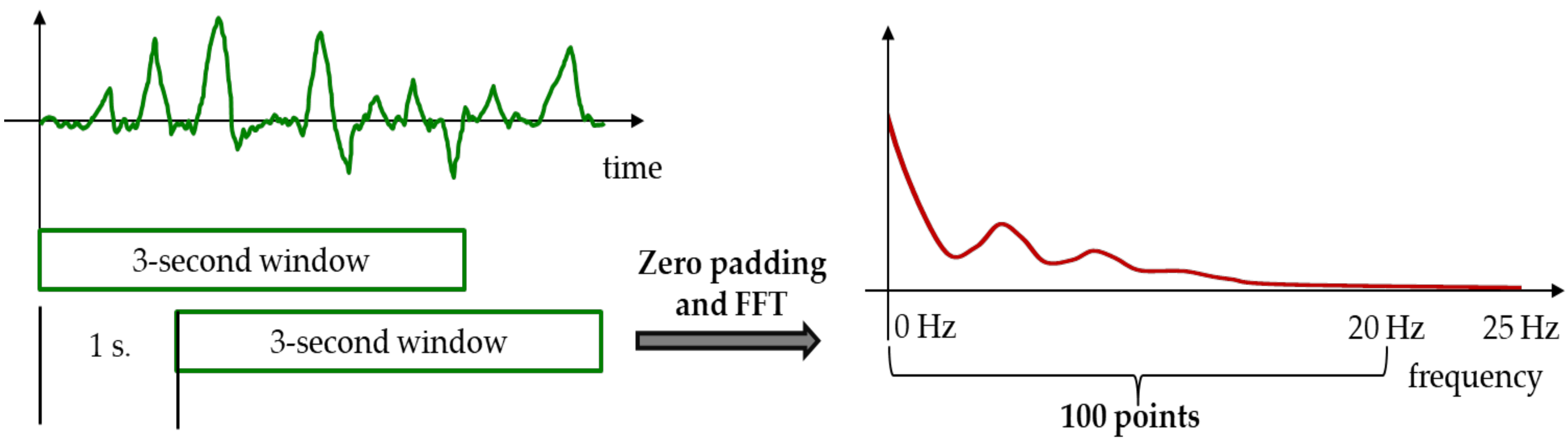
Database description

- MotionSense database
  - 6 physical activities, 24 subjects
  - iPhone 6S in trousers' front pocket
  - Accelerometer sampling at 50 Hz



Signal processing

- Two data formats of acceleration signals
  - Raw data
  - Fast Fourier Transform (FFT)

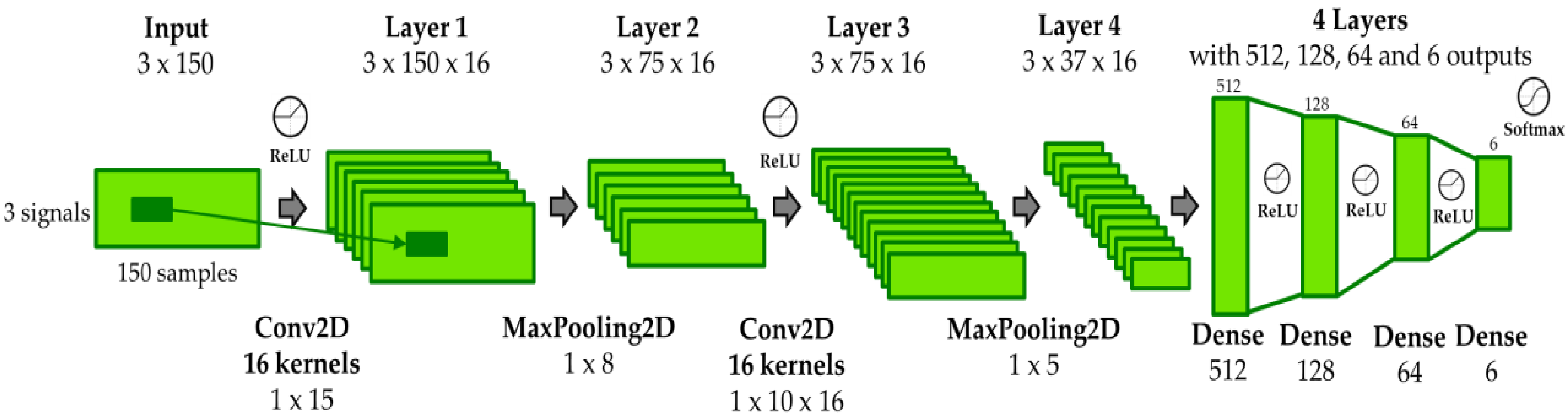


Time domain: 3-second windows of raw data

Frequency domain: spectrum of 3-second windows

Deep learning architecture

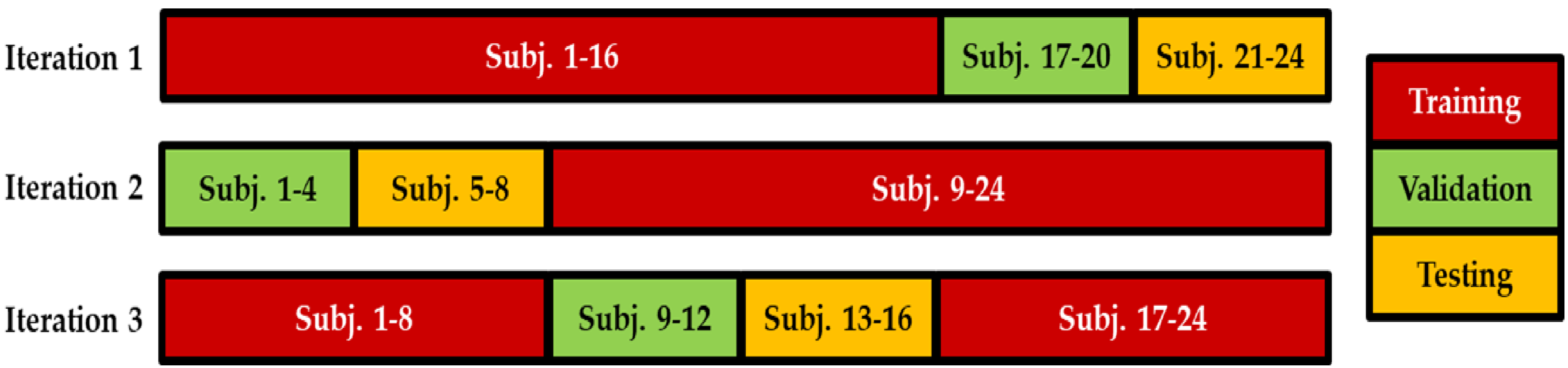
- Deep learning architecture composed of two convolutional and two max-pooling layers with decreasing kernel sizes



- Variations of this architecture
  - Number of convolutional kernels
  - Decreasing kernel sizes

Subject-Wise Cross-Validation

- Subject-wise three-fold cross-validation
  - Data from 16 subjects to training set
  - Data from 4 subjects to validation set
  - Data from 4 subjects to testing set



- Performance of testing subset with number of epochs that minimizes the error in the validation subset

Results

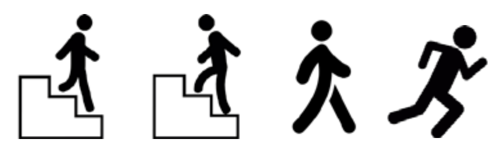
Input of deep learning architecture	Number of convolutional layers	Convolutional kernels size (x,y)	Pooling kernels size (x,y)	Accuracy (%)
Raw data	1	(1,10)	(1,5)	91.23 ± 0.55
	1	(1,15)	(1,8)	92.92 ± 0.50
	1	(1,20)	(1,10)	93.21 ± 0.49
	1	(1,30)	(1,15)	91.95 ± 0.53
	2	(1,10) / (1,10)	(1,5) / (1,5)	92.50 ± 0.51
	2	(1,15) / (1,10)	(1,8) / (1,5)	**95.28 ± 0.41**
	2	(1,15) / (1,15)	(1,8) / (1,8)	93.39 ± 0.48
	2	(1,20) / (1,10)	(1,10) / (1,5)	91.93 ± 0.53
	2	(1,20) / (1,15)	(1,10) / (1,8)	93.77 ± 0.47
	2	(1,20) / (1,20)	(1,10) / (1,10)	93.63 ± 0.47
	2	(1,30) / (1,20)	(1,15) / (1,10)	91.55 ± 0.54

Input of deep learning architecture	Number of convolutional layers	Convolutional kernels size (x,y)	Pooling kernels size (x,y)	Accuracy (%)
FFT module	1	(3,3)	(1,2)	93.34 ± 0.48
	1	(3,6)	(1,3)	**94.19 ± 0.45**
	1	(3,9)	(1,5)	93.00 ± 0.49
	2	(3,3) / (3,3)	(1,2) / (1,2)	93.13 ± 0.49
	2	(3,6) / (3,6)	(1,3) / (1,3)	92.83 ± 0.50
	2	(3,6) / (3,3)	(1,3) / (1,2)	93.39 ± 0.48

Results

- Raw data as input for classifying dynamic and static activities

- Experiments for different types of movements
  - Dynamic activities
    - ➢ 88.18% of accuracy using raw data as input
    - ➢ 90.14% of accuracy using FFT module as input



  - Static activities
    - ➢ No significant differences

Discussion

- CNNs architecture for feature extraction and activity classification better than Support Vector Machines
- Selection of deep neural network depending on the input data
- Computation of FFT equivalent to convolutional layer
- Dynamic activities characterized by information in the frequency domain

Conclusions

- HAR system based on deep learning techniques for MotionSense dataset activities
- Different transformations to find the most appropriate input data format to the deep neural network
- Two convolutional and two max-pooling layers with decreasing kernel sizes architecture using raw data as input
- FFT module as input for classifying only between dynamic activities

6th International Electronic Conference on Sensors and Applications

THANK YOU
FOR YOUR ATTENTION

Manuel Gil-Martín

ECSA-6

POLITÉCNICA

Departamento de Ingeniería Electrónica
Universidad Politécnica de Madrid