GEC 2019

VISOR Project

Human activity recognition using 3D skeletal information

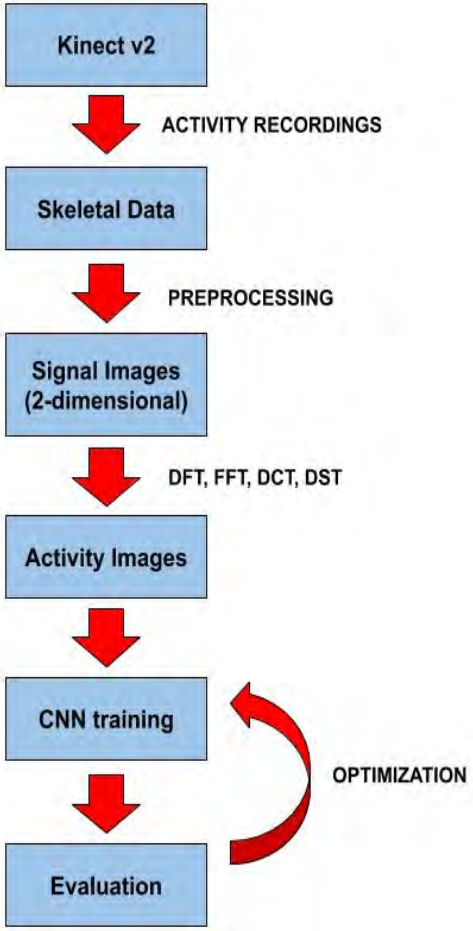
Antonios Papadakis[1], Eirini Mathe[2] , Evaggelos Spyrou [2,3]

[1]Department of Informatics and Telecommunications, University of Athens, Greece
[2]Computational Intelligence Laboratory (**CIL**), Institute of Informatics and Telecommunications, National Center for Scientific Research - “Demokritos,” Athens, Greece
[3]General Department, University of Thessaly, Lamia, Greece

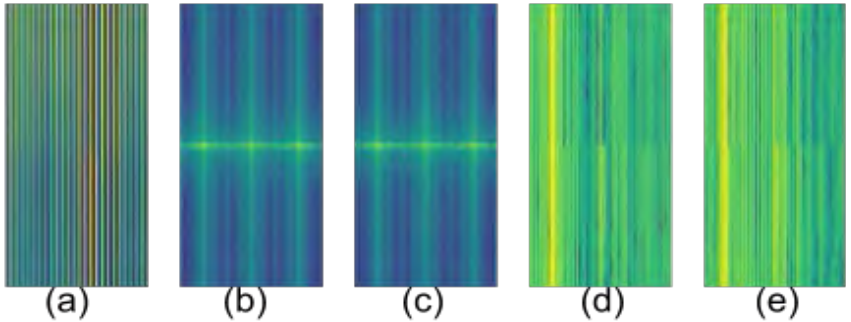
Proposed approach

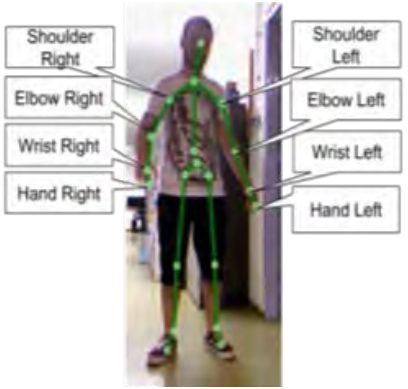
Our approach consists of:

1. Acquiring **raw** skeletal data with Kinect v2 (PKU-MMD dataset)
    - 3 camera angles (Left, Middle and Right)
1. Creation of the 2-D **signal** images
2. Transformation of the signal images into **activity** images using DFT, FFT, DCT, DST
3. Implementation and training of the proposed **CNN** architecture
4. Experimental **evaluation** and optimization of our model



Signal and Activity Images





- (a) represents a **signal** image
- (b-e) **activity** images are created by applying the following image transforms:
  - the 2-D Discrete Fourier Transform (**DFT**)
  - the 2-D Fast Fourier Transform (**FFT**)
  - the 2-D Discrete Cosine Transform (**DCT**)
  - the 2-D Discrete Sine Transform (**DST**)
- During the transformation only the **magnitude** is preserved

Experimental Results

11 selected classes of the PKU-MMD dataset

Experiment	Train	Test	DFT	FFT	DCT	DST
Single-view	M	M	**0.89**	0.84	0.83	0.86
	L	L	0.76	0.82	0.78	**0.84**
	R	R	0.84	0.85	**0.89**	0.87
Cross-view	M	L	0.62	0.61	0.63	**0.64**
	M	R	0.58	0.61	**0.65**	0.63
	L	M	0.65	0.66	0.64	**0.72**
	L	R	0.41	0.38	0.32	**0.43**
	R	M	0.56	0.59	**0.64**	0.63
	R	L	0.32	0.37	0.33	**0.39**
	M,L	R	0.60	0.60	**0.62**	**0.62**
	M,R	L	**0.60**	0.57	0.59	**0.60**
	L,R	M	0.77	0.77	0.81	**0.82**
Cross-subject	M,L,R	M,L,R	**0.85**	0.79	0.83	0.81

51 selected classes of the PKU-MMD dataset

Experiment	Train	Test	DFT	FFT	DCT	DST
Single-view	M	M	0.46	0.49	**0.66**	0.65
	L	L	0.46	0.52	0.60	**0.65**
	R	R	0.55	0.51	**0.69**	0.67
Cross-view	M	L	0.33	0.34	0.33	**0.36**
	M	R	0.32	0.31	**0.36**	0.32
	L	M	0.33	0.30	0.35	**0.39**
	L	R	**0.22**	**0.22**	0.13	0.14
	R	M	0.32	0.34	**0.36**	0.34
	R	L	**0.20**	0.19	0.13	0.14
	M,L	R	0.33	**0.34**	0.33	**0.34**
	M,R	L	0.32	0.33	0.32	**0.34**
	L,R	M	0.44	0.44	**0.55**	0.52
Cross-subject	M,L,R	M,L,R	0.50	0.52	**0.64**	0.63

- M, L and R denote the middle, left and right camera angles
- **0.89** and **0.69** represent the best accuracy results of our model for the 11 and 51 classes scenarios respectively. Both results appeared in the single view test cases.

Thank you