Robust multi-dimensional motion features for first-person vision activity recognition

Published in: Computer Vision and Image Understanding (CVIU), 2016

Authors: Girmaw Abebe, Andrea Cavallaro, Xavier Parra

Centre for Intelligent Sensing
Queen Mary University of London

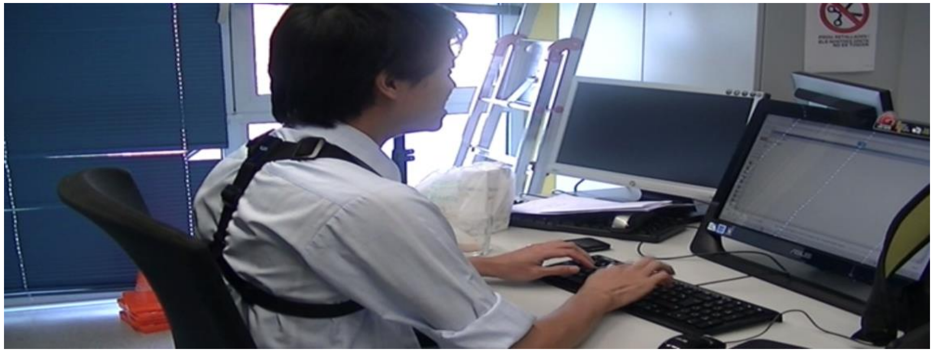
Introduction

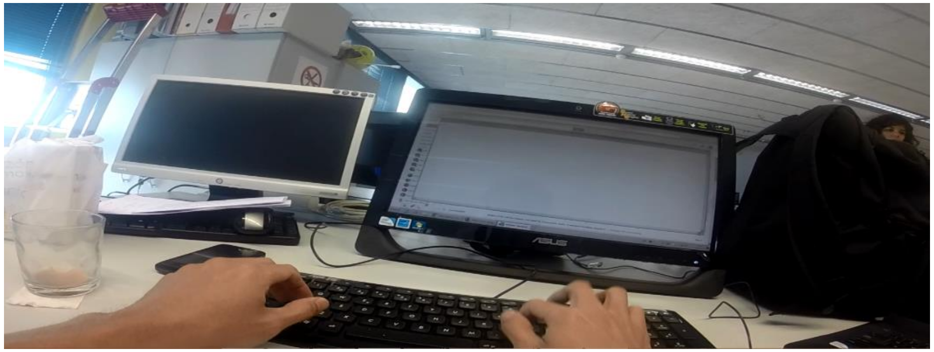
Third-person vision

Sensing is external

First-person vision

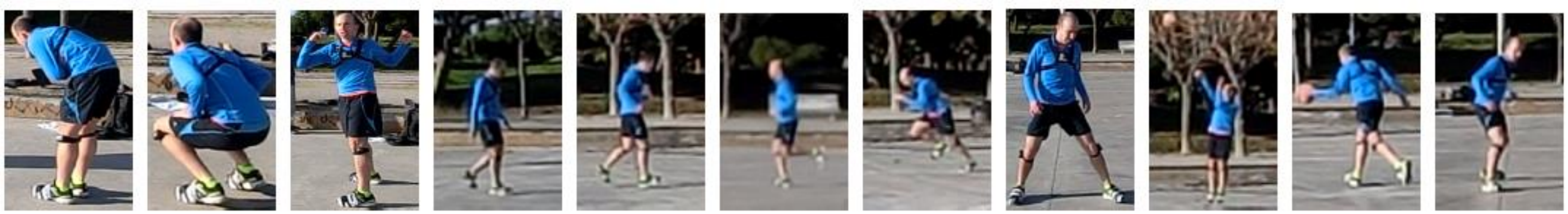
Sensing is ego-centric





Problem definition

- Given first-person videos of ambulatory activities, we develop a robust motion feature in order to recognize the activities
- Ambulatory activities involve full-body motions

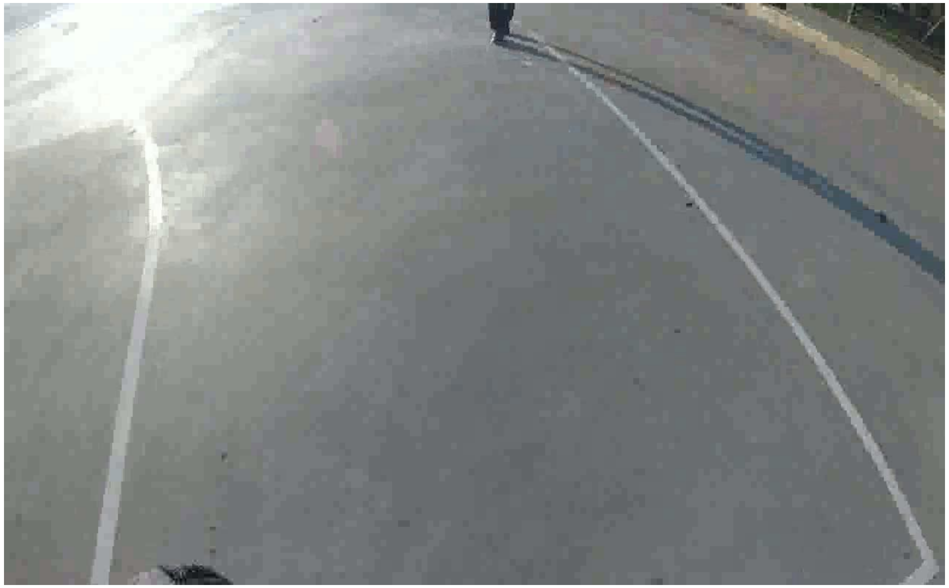


Third-person view

First-person view

Challenges

- Complex ego-motion
- Motion parallax and blur
  - E.g., Dribble and Sprint
- Local motions
  - E.g., Appearance of people
- Mounting point variations
  - E.g., Chest vs Head
- Limited datasets



Related work: acquisition

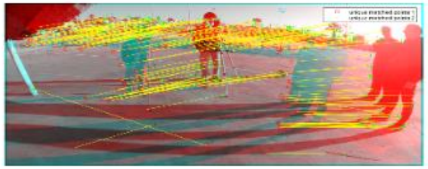
- Acquisition device
  - Lab prototypes [zhang2010,2011, Zhan2012,2014,2015, Nam2013]
  - Commercial cameras [Ryoo2015, Kitani2011, Poleg2016, Poleg2014]
- Mounting positions
  - Head mounts [Kitani2011, Poleg2016, Poleg2014]
  - Chest [zhang2010,2011]
  - Wrist [Nam2013]
- Preprocessing
  - Data resizing
  - Filtering

[zhang2010,2011]

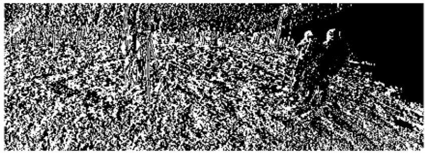
[Poleg2014]

Related work: feature extraction

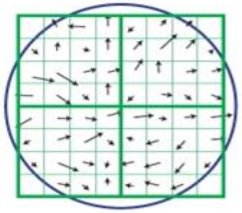
- Keypoint-based [zhang2010,2011]



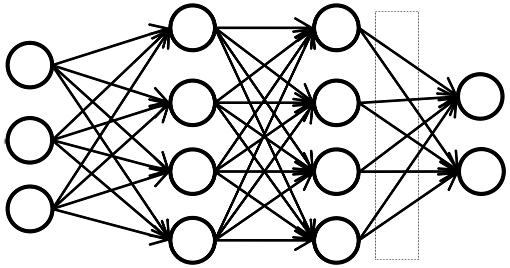
- Optical flow-based
  - Grid optical flow [Zhan2012,2014, Nam2013, Poleg2014
  - Magnitude [Kitani2011]
  - Direction [Ryoo2013,2015, Iwashita2014]
  - Frequency domain analysis [Kitani2011]



- Intra-frame appearance [Kitani2014,Ryoo2015]

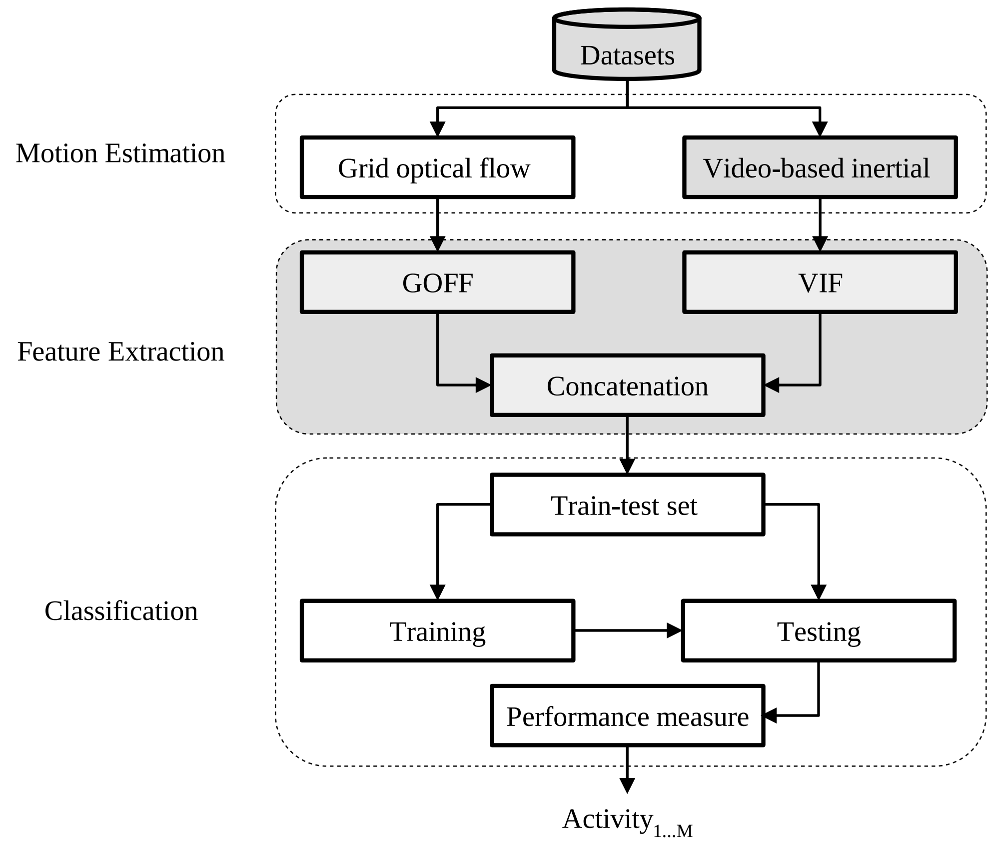


[Lowe1999]



Proposed approach

- Encode significant motion variations
- Generate virtual inertial features
- Multiple validation

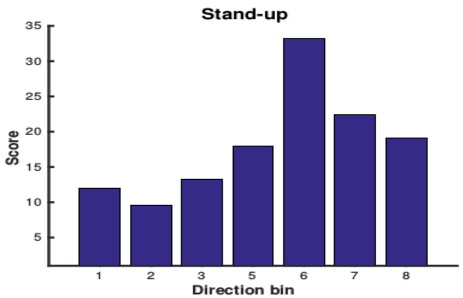


- Contributions are highlighted

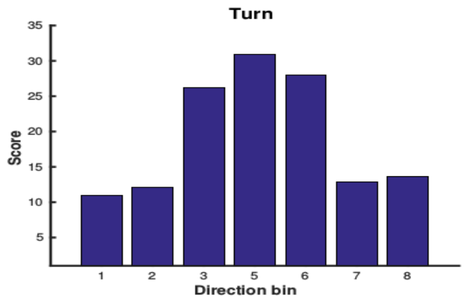
Grid Features: examples

- Proposed features are shown to discriminate activities.
  - E.g., MDHF: motion-direction histogram feature

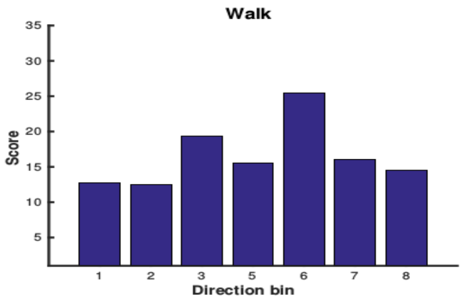
(a)



(b)



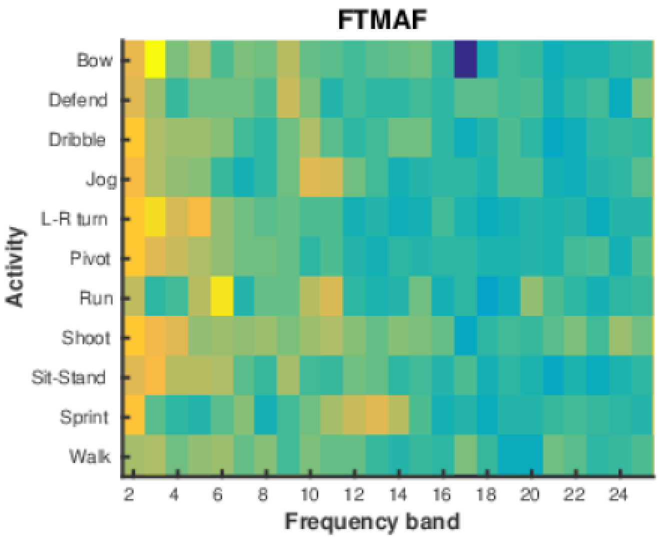
(c)

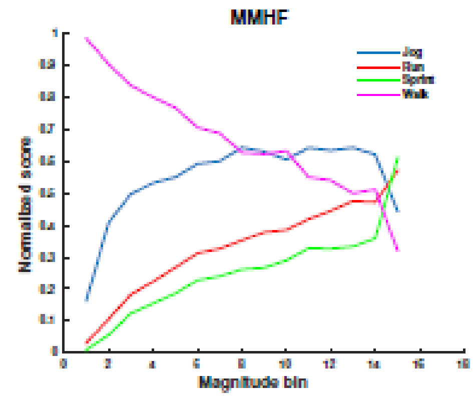


(d)

Grid Features: more examples

- Frequency feature of motion direction (FTMAF)





- Motion magnitude histogram (FTMAF)

Evaluation

- Datasets
  - IAR

  - BAR

  - JPL [Ryoo2013]

  - DogC [Iwashita2014]

Results on our datasets

Dataset	Method	A	P	R	S	F_1	KNN
IAR	AP [zhan2015]	92	85	42	**99**	56	49
	MRGF [zhang2011]	97	90	**87**	98	**88**	**79**
	MBH [Kitani2011]	91	62	68	94	65	67
	RMF (Proposed)	**97**	**91**	85	**99**	**88**	78
BAR	AP [zhan2015]	90	24	14	97	18	31
	MRGF [zhang2011]	89	35	39	93	37	48
	MBH [Kitani2011]	95	63	67	97	64	71
	RMF (Proposed)	**98**	**81**	**79**	**99**	**80**	**78**

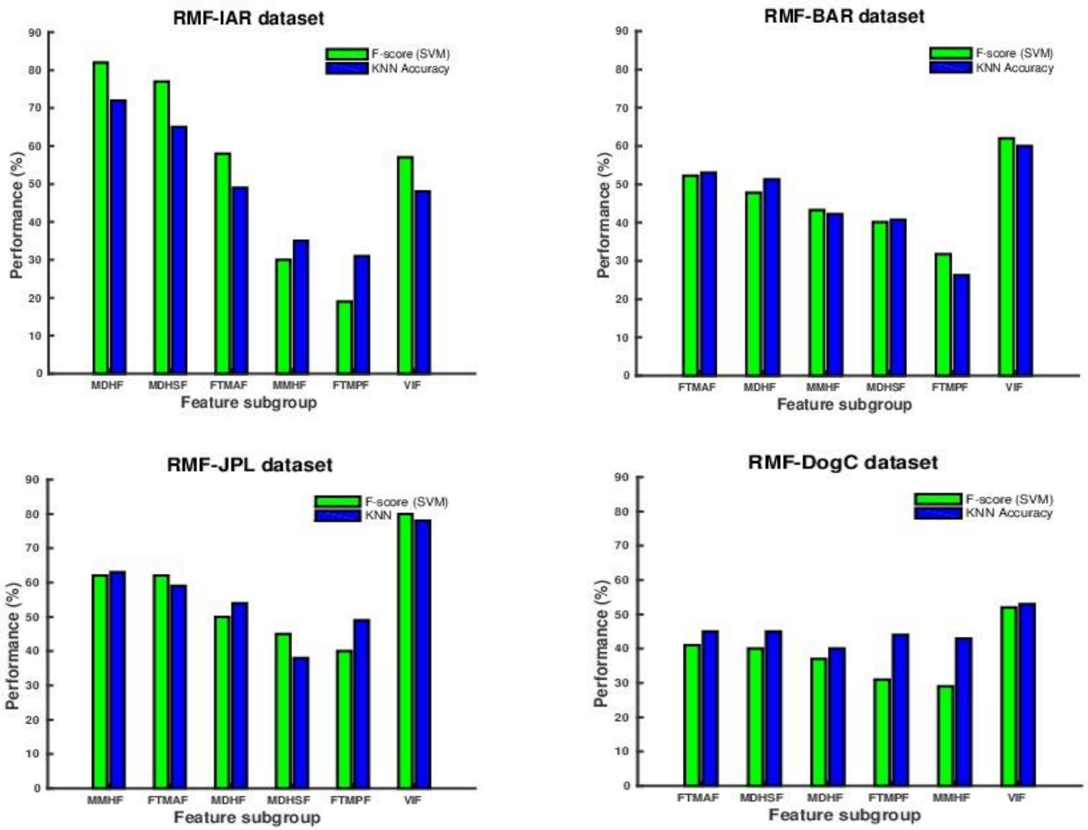
Key:
A: Accuracy
P: Precision
R: Recall
S: Specificity
F_1: F-score
KNN: Recall outpu
of KNN classifier

Results on public datasets

Dataset	Method	A	P	R	S	F_1	KNN
JPL	AP [zhan2015]	76	5	16	86	7	34
	MRGF [zhang2011]	85	55	72	87	62	55
	MBH [Kitani2011]	87	66	53	92	59	61
	RMF (Proposed)	**96**	**87**	**85**	**97**	**86**	**82**
DogC	AP [zhan2015]	87	39	30	92	34	47
	MRGF [zhang2011]	88	39	39	94	39	42
	MBH [Kitani2011]	86	38	27	92	32	51
	RMF (Proposed)	**92**	**62**	**59**	**96**	**61**	**58**

Key:
A: Accuracy
P: Precision
R: Recall
S: Specificity
F_1: F-score
KNN: Recall outpu
of KNN classifier

Discussion: features



Summary

- Multi-dimensional motion features encode direction, magnitude and dynamics
- Inertial features generated from video proven to be useful
- Collection of new publicly available datasets
- Limitations: small datasets, mounting positions (self-occlusions)

Thank you!

Questions?