



# Center for Research in Computer Vision

UNIVERSITY OF CENTRAL FLORIDA

## FINAL ORAL EXAMINATION

*OF*

**SHERVIN ARDESHIR**

B.S., SHARIF UNIVERSITY OF TECHNOLOGY, 2013  
M.S., UNIVERSITY OF CENTRAL FLORIDA, 2016

*FOR THE DEGREE OF*

**DOCTOR OF PHILOSOPHY**  
(COMPUTER SCIENCE)

27 June, 2018, 2:00 P.M.  
HEC 110

### **DISSERTATION COMMITTEE**

Professor Ali Borji, *Chairman*, [aborji@crcv.ucf.edu](mailto:aborji@crcv.ucf.edu)

Professor Mubarak Shah, *Vice-chair*, [shah@crcv.ucf.edu](mailto:shah@crcv.ucf.edu)

Professor Haiyan Hu, [haihu@cs.ucf.edu](mailto:haihu@cs.ucf.edu)

Professor George Atia, [george.atia@ucf.edu](mailto:george.atia@ucf.edu)

# DISSERTATION RESEARCH IMPACT

Relating first-person (egocentric) and third-person (exocentric) videos is challenging due to the drastic difference in camera motion and viewpoint. Exploring this relationship could have many applications in surveillance networks, sport analysis, etc. Also, since third-person vision has a longer history in computer vision, successfully transferring knowledge across the two views could lead to progress in performing computer vision tasks in the first-person domain.

This dissertation contributes to studying this relationship by exploring:

(1) self-identification of an egocentric camera holder in a top-view exocentric video, (2) temporally aligning first-person and top-view third-person videos (3) jointly solving self-identification, and re-identification across the two views, and (4) performing action classification across egocentric and exocentric views. We conclude that many of the aforementioned problems are highly interconnected, and jointly addressing them could lead to a better understanding the relationship between these two drastically different perspectives.

## SELECTED PUBLICATIONS (h-index: 5, total citation: 76)

1. **Egocentric Meets Top-view**, S. Ardeshir and A. Borji, in *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2018
2. **An Exocentric Look at Egocentric Actions, and Vice versa**, S. Ardeshir, and A. Borji, in *Computer Vision and Image Understanding (CVIU)*, 2018.
3. **Ego2Top: Matching Viewers in Egocentric and Top-view Videos**, S. Ardeshir, and A. Borji, in *European Conference on Computer Vision (ECCV)*, 2016.
4. **Geo-semantic Segmentation**, S. Ardeshir, K.M. Collins-Sibley<sub>2</sub> and M. Shah, in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
5. **GIS-assisted Object Detection and Geospatial Localization**, S. Ardeshir, A.R. Zamir, A. Torroella, and M. Shah, in *European Conference on Computer Vision (ECCV)*, 2014.
6. **GPS-tag Refinement using Random-walks with an Adaptive Damping Factor**, A. R. Zamir, S. Ardeshir and M. Shah, in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

# DISSERTATION

## RELATING FIRST-PERSON AND THIRD-PERSON VISION

Thanks to the availability and increasing popularity of wearable devices such as GoPro cameras, smart phones and glasses, we have access to a plethora of videos captured from the first person perspective. Capturing the world from the perspective of one's self, egocentric videos bear characteristics distinct from the more traditional third-person (exocentric) videos. First and third person videos have been studied in the past in the computer vision community. However, the relationship between first and third person vision has yet to be fully explored. This thesis explores this relationship in several aspects. We explore supervised and unsupervised approaches for relating these two views seeking different objectives such as identification, temporal alignment, and action classification.

In chapter 3, we make the first attempt in this direction by performing identification across egocentric and top-view videos. Having a set of egocentric videos and a top-view video, we identify each individual egocentric viewer in the top-view video. We model each view (egocentric or top) using a graph, and use spectral graph matching techniques in order to attain correspondences across the two views. Even though we obtain promising assignment results in many cases, we observe that this problem can become extremely challenging when videos are not temporally aligned. Hence, in chapter 4 we propose approaches taking into account the time-delays in the spectral graph matching formulation. We evaluate our methods in terms of ranking and assigning egocentric viewers to identities present in the top-view videos. We evaluate the effectiveness of our framework on a dataset of 50 top-view and 188 egocentric videos captured under different conditions.

In chapter 5 we address a shortcoming in the self-identification task. We show that previous methods are highly dependent on the completeness of the egocentric set and thus, are not suitable for scenarios in which there is only one egocentric and one top-view video available. Given such scenario, we identify the egocentric camera holder in the top-view video (self-identification), identify the humans visible in the content of the egocentric video in the top-view video (re-identification), and temporally align the two videos. We show that each of these tasks is highly dependent on the other two. We propose a unified framework to jointly address all three problems.

Another task explored in this thesis is relating action information across first-person (egocentric) and third-person (exocentric) views. In chapter 6, we investigate two different, yet highly interconnected problems including cross-view action classification and action based video retrieval. We perform action classification in one domain using the knowledge transferred from the other domain. Also, given a video in one view, we retrieve videos from the same action class in the other view. We collect a new cross-domain dataset of egocentric-exocentric action videos containing 14 action classes and 3569 videos (1676 collected egocentric videos and 1893 exocentric videos borrowed from UCF 101). Our results demonstrate the possibility of transferring action information across the two domains and performing action based matching and retrieval in seen and unseen classes.



## **SHERVIN ARDESHIR**

1989	Born in Tehran, Iran
2013	B.Sc., Sharif University of Technology, Tehran, Iran.
2016	M.Sc., University of Central Florida, Orlando, Florida.
2016	Robotics Intern, iRobot, Pasadena, California.
2013-18	Ph.D., University of Central Florida, Orlando, Florida.

## **SELECTED AWARDS**

2017	CVPR Doctoral Consortium Travel Award
------	---------------------------------------