Center for Research in Computer Vision

UNIVERSITY OF CENTRAL FLORIDA

FINAL ORAL EXAMINATION

OF

KHURRAM SOOMRO

B.Sc, LAHORE UNIVERSITY OF MANAGEMENT SCIENCES, 2007
M.Sc, LAHORE UNIVERSITY OF MANAGEMENT SCIENCES, 2011

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

(COMPUTER SCIENCE)

03 November, 2017, 2:00 PM.
HEC 101

DISSERTATION COMMITTEE

Professor Mubarak Shah, *Chairman, shah@crcv.ucf.edu*
Professor Mark Heinrich, *heinrich@cs.ucf.edu*
Professor Haiyan Hu, *haihu@cs.ucf.edu*
Professor Ulas Bagci, *bagci@crcv.ucf.edu*
Professor Hae-Bum Yun, *hae-bum.yun@ucf.edu*

DISSERTATION RESEARCH IMPACT

Recognizing and localizing actions has been fundamental to video understanding in computer vision. It is a challenging problem, which has a wide variety of applications from monitoring and security in surveillance videos, to video search, action retrieval, multimedia event recounting and human-computer interaction. This dissertation contributes by proposing: (1) an efficient approach for action localization, which is scalable to real-world applications having videos of higher resolution and longer duration; (2) an online localization method that can anticipate actions/ interactions and predict them in a timely manner; and (3) an unsupervised action localization approach that can automatically discover and localize actions, without the need of manually labeled and annotated training videos. Moreover, real-time applications can monitor the elderly to alert the care giver, detect abnormal actions of criminal nature or timely detection of human actions for autonomous driving.

SELECTED PUBLICATIONS (h-index: 7, total citations: 839)

1. **Online Localization and Prediction of Actions and Interactions.** Khurram Soomro, Haroon Idrees and Mubarak Shah, *IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**)*, 2017.

2. **Unsupervised Action Discovery and Localization in Videos.** Khurram Soomro and Mubarak Shah, *IEEE International Conference on Computer Vision (**ICCV**)*, 2017.

3. **Predicting the Where and What of Actors and Actions through Online Action Localization.** Khurram Soomro, Haroon Idrees and Mubarak Shah, *IEEE International Conference on Computer Vision and Pattern Recognition (**CVPR**)*, 2016.

4. **Action Localization in Videos through Context Walk.** Khurram Soomro, Haroon Idrees and Mubarak Shah, *IEEE International Conference on Computer Vision (**ICCV**)*, 2015.

5. **Tracking when the camera looks away.** Khurram Soomro, Salman Khokhar and Mubarak Shah, *IEEE International Conference on Computer Vision Workshop (**ICCVW**)*, 2015.

6. **Detecting Humans in Dense Crowds using Locally-Consistent Scale Prior and Global Occlusion Reasoning.** Haroon Idrees, Khurram Soomro and Mubarak Shah, *IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**)*, 2015.

7. **Action Recognition in Realistic Sports Videos.** Khurram Soomro and Amir R. Zamir, *Computer Vision in Sports, Springer International Publishing*, 2014.

8. **UCF101: A Dataset of 101 Human Actions Classes from Videos in the Wild.** Khurram Soomro, Amir R. Zamir and Mubarak Shah, CRCV-TR-12-01, 2012.

PATENT

1. **Classification of barcode tag conditions from top view sample tube images for laboratory automation.** Khurram Soomro, Y. J. Chang, S. Kluckner, W. Wu, B. Pollack and T. Chen. **US Patent: WO2016133915 A1.**

DISSERTATION

ONLINE, SUPERVISED AND UNSUPERVISED ACTION LOCALIZATION IN VIDEOS

Action recognition involves classification of a given video in terms of a set of action labels, whereas action localization determines the location of an action in addition to its class. Many of the existing action localization approaches exhaustively search (spatially and temporally) for an action in a video. However, as the search space increases with high resolution and longer duration videos, it becomes impractical to use such sliding window techniques. The first part of this dissertation presents an efficient approach for localizing actions by learning contextual relations in training, in the form of relative locations between different video regions (supervoxels). These relations are captured as displacements from all the supervoxels in a video to those belonging to foreground actions. Then, given a testing video, we select a supervoxel randomly and use the context information acquired during training to estimate the probability of each supervoxel belonging to the foreground action. The walk proceeds to a new supervoxel and the process is repeated for a few steps. A Conditional Random Field (CRF) is then used to localize actions, whose confidences are obtained using SVMs.

In the above method and typical approaches to this problem, localization is performed in an offline manner where all the frames in the video are processed together. This prevents timely localization and prediction of actions/interactions - an important consideration for many tasks including surveillance and human-machine interaction. Therefore, in the second part of this dissertation we propose an online approach to the challenging problem of localization and prediction of actions/interactions in videos. In this approach, we estimate human poses at each frame and train discriminative appearance models using the superpixels inside the pose bounding boxes. Since the pose estimation per frame is inherently noisy, the conditional probability of pose hypotheses at current time-step (frame) is computed using pose estimations in the present frame and their consistency with poses in the previous frames. Next, both the superpixel and pose-based foreground likelihoods are used to infer the location of actors at each time through CRF. For online prediction of action/interaction confidences, we propose an approach based on Structural SVM that is trained with the objective that confidence of an action/interaction increases as time progresses.

Above two approaches rely on human supervision in the form of assigning action class labels to videos and annotating actor bounding boxes in each frame of training videos. Therefore, in the third part of this dissertation we address the problem of unsupervised action localization. Given unlabeled data without annotations, this approach aims at: 1) Discovering action classes and 2) Localizing actions in videos. It begins by applying spectral clustering on a set of unlabeled training videos. For each cluster, an undirected graph is constructed to extract a dominant set. Next, a discriminative clustering approach is applied by training a classifier for each cluster, to iteratively select videos from the non-dominant set and obtain complete video action classes. Annotations for training videos are obtained by over-segmenting videos into supervoxels and constructing a directed graph to apply a variant of knapsack problem. Knapsack selects supervoxels to generate action annotations for each video. These annotations and discovered action classes are used to train our action classifier. During testing, actions are localized using Knapsack approach, and SVM is used to recognize these actions.

KHURRAM SOOMRO

1984	Born in Larkana, Pakistan
2003-07	B.Sc., Lahore University of Management Sciences, Pakistan
2007-09	Analyst Software Engineer, The Resource Group, Pakistan
2009-11	M.Sc., Lahore University of Management Sciences, Pakistan
2014	Computer Vision Intern, Siemens, Princeton, NJ
2011-17	Ph.D., University of Central Florida, Orlando, FL

SELECTED AWARDS

2015	Gerald R. Langston Endowed Scholarship
2015	ICCV 2015 Doctoral Consortium Award
2016	CVPR 2016 Doctoral Consortium Award
2016	UCF Graduate Research Forum Winner
2016	Statewide Graduate Student Research Symposium Winner

INVITED TALK

2017	*Action Localization in Videos*, Department of Computer Sciences and Cybersecurity, School of Computing, **Florida Institute of Technology**