



WAQAS SULTANI

1985 Born in Gujranwala, Pakistan
2006 B.Sc.(Hons), University of Engineering and Technology,
Taxila, Pakistan
2010 M.Sc , Seoul National University, South Korea
2010-11 Lecturer, Comsats Institute of Technology, Islamabad
2014 Computer Vision Intern, Xerox, Webster, NY
2011-17 Ph.D., University of Central Florida, Orlando, Florida.

SELECTED AWARDS

2008 MS scholarship by Higher Education Commission, Pakistan
2014 IT Performance Travel Grant for CVPR 2014.



Center for Research in Computer Vision

UNIVERSITY OF CENTRAL FLORIDA

FINAL ORAL EXAMINATION

OF

WAQAS SULTANI

B.Sc, UNIVERSITY OF ENG. & TECH, TAXILA, 2006
M.Sc, SEOUL NATIONAL UNIVERSITY, 2010

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY (COMPUTER SCIENCE)

07 July, 2017, 4:00 PM.
HEC 110

DISSERTATION COMMITTEE

Professor Mubarak Shah, *Chairman, shah@crcv.ucf.edu*
Professor Ulas Bagci, *bagci@crcv.ucf.edu*
Professor Guo Jun Qi, *guojun.qi@ucf.edu*
Professor Hae-Bum Yun, *hae-bum.yun@ucf.edu*

DISSERTATION RESEARCH IMPACT

This dissertation contributes to human action recognition and localization in real-world environments. The development of a robust action recognition system has several practical applications and can have the significant impact on our daily lives. Some of its applications are: (1) It can enhance human-computer interaction and let robots do their tasks automatically by properly responding to human actions; (2) In surveillance applications, it can automatically detect suspicious and abnormal activities such as crimes and illegal activities and has potential to save several lives; (3) It can facilitate detailed analysis of sports videos and can help athletes to further improve their skills; (4) It can alleviate some of the problems in text-based video indexing and retrieval and improve them through recognizing activities within videos; (5) finally, it can contribute in developing intelligent environment systems to facilitate taking care of elderly people.

SELECTED PUBLICATIONS AND PATENTS

Automatic Action Annotation in Weakly Labeled Videos. Waqas Sultani and Mubarak Shah, Computer Vision and Image Understanding (CVIU), 2017.

Unsupervised Action Proposal Ranking through Proposal Recombination. Waqas Sultani, Dong Zhang and Mubarak Shah, Computer Vision and Image Understanding (CVIU), 2017.

What if we do not have multiple videos of the same action? Video Action localization using Web Images . Waqas Sultani and Mubarak Shah, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

Human action recognition across dataset using foreground focused Histogram decomposition. Waqas Sultani and Imran Saleemi, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

System and Method for Detecting Settle Down Time using Computer Vision Techniques. Robert P. Loce, Waqas Sultani, Beilei Xu, Hao Wu. US Patent: 9,378,421 B2.

System and Method for Seat Occupancy Detection from Ceiling Mounted Camera using Robust Adaptive Threshold Criteria., R . Loce, B. Xu, F. Wade, M. Ann, P.Wall, Waqas Sultani, H. Wu, E.H US Patent: 9,384, 396, B2.

Abnormal Traffic Detection using Intelligent Driver Model, Waqas Sultani, and Jin Young Choi, IEEE International Conference on Pattern Recognition (ICPR), 2010.

DISSERTATION

WEAKLY LABELED ACTION RECOGNITION AND DETECTION

Research in human action recognition strives to develop increasingly generalized methods that are robust to intra-class variability and inter-class ambiguity. Training such models need thousands of precise spatio-temporal manual annotations, which require many human annotators, hundreds of hours, and are subject to human biases.

In the first part of this dissertation, we explore the reasons for poor classifier performance when tested on novel datasets, and quantify the effect of scene background on action representations and recognition. We propose a new method to obtain a measure of confidence in each pixel of the video being a foreground region using motion, appearance, and saliency together in a 3D-Markov Random Field (MRF) based framework. We also propose multiple ways to exploit the foreground confidence: to improve bag-of-words vocabulary, histogram representation of a video, and a novel histogram decomposition based representation and kernel.

The above-mentioned method does not provide the precise spatio-temporal location of the actor and needs manual spatio-temporal annotations to train an action detector. Therefore, in the second part of this dissertation, we propose a weakly labeled approach to automatically obtain spatio-temporal annotations of actors in action videos. We first obtain a large number of action proposals in each video. To capture a few most representative action proposals in each video, we rank them using motion and saliency cues and select a few proposals using MAP based proposal subset selection method. Our next challenge is to iteratively select one proposal from each video, so that all proposals are globally consistent. We formulate this as Generalized Maximum Clique Graph problem. The output of our method is the most action representative proposals from each video which are used to train action detector.

The above-mentioned annotation method uses multiple videos of the same action. Therefore, in the third part of this dissertation, we tackle the problem of spatio-temporal action localization in a video, without assuming the availability of multiple videos or any prior annotations. We present a novel approach for action localization in videos using web images. Given a video, first we generate multiple spatio-temporal action proposals. To obtain the most action representative proposals, we reconstruct action proposals in the video by leveraging the action proposals in images. We solve this optimization problem using the variant of two-metric projection algorithm.

Finally, we propose a framework to generate a few better action proposals that are ranked properly. We first divide each action proposal into sub-proposals and then use Dynamic Programming based graph optimization scheme to select the optimal combinations of sub-proposals from different proposals and assign each new proposal a score. We propose a new unsupervised image-based actionness detector that leverages web images and employs it as one of the node scores in our graph formulation. We demonstrate that properly ranked proposals produce significantly better action detection as compared to state-of-the-art proposal based methods.

Our extensive experimental results on different challenging and realistic action datasets, comparisons with several competitive baselines and detailed analysis of each step of proposed methods validate the proposed ideas and frameworks.