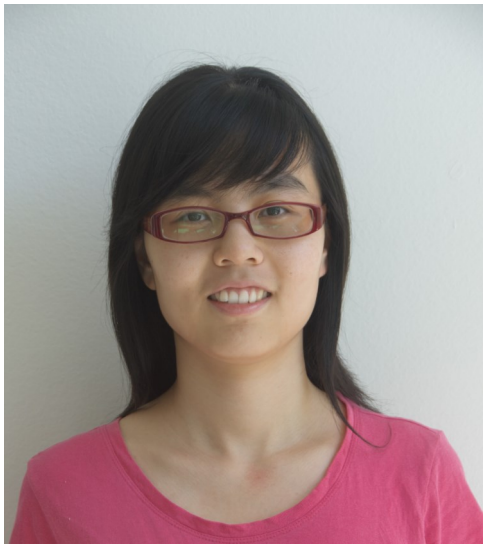

YICONG TIAN

1989	Born in Luoyang, China
2011	B.S., Beijing Univ. of Posts and Telecom., Beijing, China
2015	Software Engineer Intern, Google, Mountain View, CA
2011-18	Ph.D., University of Central Florida, Orlando, FL
2017-	Software Engineer, Google, Mountain View, CA

UCF | Center for Research in Computer Vision

UNIVERSITY OF CENTRAL FLORIDA

FINAL ORAL EXAMINATION

OF

YICONG TIAN

B.S., Beijing University of Posts and Telecommunications, 2011

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

(COMPUTER SCIENCE)

25 October, 2018, 12:30 P.M.
CREOL 103

DISSERTATION COMMITTEE

Professor Mubarak Shah, *Chair*, *shah@crcv.ucf.edu*
Professor Ulas Bagci, *bagci@crcv.ucf.edu*
Professor Fei Liu, *feiliu@cs.ucf.edu*
Professor John Walker, *John.Walker@ucf.edu*

DISSERTATION RESEARCH IMPACT

Human action detection, human tracking and segmentation are fundamental tasks in computer vision. They have a variety of applications. Some examples are: (1) They facilitate automatic sports video analysis, helping coaches and athletes analyze their movements and improve their skills effectively; (2) They can be used to detect abnormal activities and track suspicious persons in surveillance videos, playing an important role in public safety; (3) They are key components in self-driving car, where pedestrians are detected and tracked so self-driving car could make safe moves.

This dissertation makes contributions to these tasks by proposing: (1) A spatiotemporal deformable part model for action detection that is robust to intra-class variation and clutter; (2) A tracker that combines detection and data association in one framework, it helps resolve the ambiguities caused by occlusion and target articulation; (3) A framework that couples target tracking with segmentation and outputs pixel-wise target labels, it helps resolve difficulties from occlusion, ID-switch and track drifting.

SELECTED PUBLICATIONS (total citation: 282)

1. **On Detection, Data Association and Segmentation for Multi-target Tracking,** Yicong Tian, Afshin Dehghan and Mubarak Shah, *in IEEE Transactions on Pattern Analysis and Machine Intelligence (**PAMI**), 2018*
2. **Cross-View Image Matching for Geo-localization in Urban Environments,** Yicong Tian, Chen Chen and Mubarak Shah, *in IEEE International Conference on Computer Vision and Pattern Recognition (**CVPR**),* 2017.
3. **Target Identity-aware Network Flow for Online Multiple Target Tracking,** Afshin Dehghan, Yicong Tian, Philip. H. S. Torr and Mubarak Shah, in *IEEE International Conference on Computer Vision and Pattern Recognition (**CVPR**),* 2015.
4. **Spatiotemporal Deformable Part Models for Action Detection,** Yicong Tian, Rahul Sukthankar and Mubarak Shah, *in IEEE International Conference on Computer Vision and Pattern Recognition (**CVPR**),* 2013.

DISSERTATION

HUMAN ACTION DETECTION, TRACKING AND SEGMENTATION IN VIDEOS

This dissertation addresses the problem of human action detection, human tracking and segmentation in videos. They are fundamental tasks in computer vision and are extremely challenging to solve in realistic videos. For action detection, the challenges include intra-class variation, camera motion and cluttered scenes. Occlusion, target interactions and articulations pose difficulty in multiple human tracking and segmentation. In this dissertation novel methods for action detection, human tracking and segmentation in videos are proposed.

Firstly, we propose a novel approach for action detection by exploring the generalization of deformable part models from 2D images to 3D spatiotemporal volumes. Actions are treated as spatiotemporal patterns and a deformable part model is generated for each action from a collection of examples. For each action model, the most discriminative 3D subvolumes are automatically selected as parts and the spatiotemporal relations between their locations are learned. By focusing on the most distinctive parts of each action, our models adapt to intra-class variation and show robustness to clutter.

The above approach deals with detecting action performed by a single person. When there are multiple humans in the scene, first humans need to be segmented and tracked from frame to frame before action detection and recognition can be performed. Next, we propose a novel approach for multiple object tracking (MOT) by formulating detection and data association in one framework. Our method allows us to overcome the confinements of data association based MOT approaches; where the performance is dependent on the object detection results provided at input level. At the core of our method lies structured learning which learns a model for each target and infers the best location of all targets simultaneously in a video clip. The inference of our structured learning is done through a new Target Identity -aware Network Flow (TINF), where each node in the network encodes the probability of each target identity belonging to that node. We show that automatically detecting and tracking targets in a single framework can help resolve the ambiguities due to frequent occlusion and heavy articulation of targets.

In the tracking method described above targets are represented by bounding boxes, which is a coarse representation. However, pixel-wise object segmentation provides fine level segmentation of targets, which is desirable for later tasks. Finally, we propose a tracker that simultaneously solves three main problems: detection, data association and segmentation. This is especially important because the output of each of those three problems are highly correlated and the solution of one can greatly help improve the others. The proposed algorithm consists of two main components: structured learning and Lagrange dual decomposition. The first component comes from TINF tracker proposed above. The second component is Lagrange dual decomposition, which combines TINF tracker with a segmentation algorithm. The proposed approach achieves more accurate segmentation results and also helps better resolve typical difficulties in multiple target tracking, such as occlusion, ID-switch and track drifting.