STEVENS
INSTITUTE OF TECHNOLOGY
1870

Ph.D. DISSERTATION DEFENSE

Candidate:	Eric C. Joyce
Degree:	Doctor of Philosophy
School/Department:	Charles V. Schaefer, Jr. School of Engineering and Science / Computer Science
Date:	Tuesday, December 9th, 2025
Time/Location:	1:30 PM, Gateway North 303
Title:	Balancing Decision and Discretion in Temporal Action Localization, Object Pose Estimation, and Robotic Grasping
Chairperson:	Dr. Philippos Mordohai, Department of Computer Science, School of Engineering & Sciences
Committee Members:	Dr. Enrique Dunn, Department of Computer Science, School of Engineering & Science Dr. Jonggi Hong, Department of Computer Science, School of Engineering & Science Dr. Ioannis Stamos, Department of Computer Science, Hunter College, City University of New York

ABSTRACT

Objects perceived in three-dimensional space serve as visual evidence, according to which downstream agents take action in the real world. However, since any vision-system may miss or hallucinate detections, we would like to obtain a measure of perception confidence to determine whether to act or not. For the task of temporal action localization (TAL), the main challenges are high intra-class variability and a large, diverse background class. In online mode, these factors make it difficult to decide when to commit to a particular action label. We developed a real-time TAL system for a VR training system designed for employees operating in an electrical substation. Our system requires a small amount of data to perform in egocentric RGB-D video streams. We address TAL's challenges by using a flexible frame descriptor based largely on objects detected in the 2D egocentric perception, dynamic time warping, and a novel approach to database construction. We demonstrate how a tradeoff between accuracy and speed may be managed by decimating the background class.

Turning from AR/VR to robotic grasping, it is possible to make a six-degree-of-freedom (6-DoF) estimate about the pose of a known, rigid object visible in a single RGB image. A robotic hand may attempt to grasp the object based on this estimate. However, factors like occlusion and clutter can introduce pose-estimate errors sufficiently large to cause a grasp attempt to fail. We present a method for predicting the success of a robotic grasp before the grasp is attempted. This allows a real-world agent to first judge whether it should act, given its input. We complement our study of grasp success with a more general evaluation of RGB-based 6-DoF object pose estimators. Besides

gaining intuition about how accurate these estimators are given only what an autonomous agent sees, we are interested in the degree to which they can serve as the sole perception mechanism for robotic grasping.

Finally, we propose a promising direction for a deterministic uncertainty model (DUM) that will predict 6-DoF poses for known objects perceived in an RGB image and express uncertainty about its own predictions in a single forward pass. The ability to predict uncertainty has tremendous potential for improving operations in robotics, autonomous navigation, and AR.