



Yang Yang

1985 Born in Beijing, China
2004-08 B.S., Beijing University of Technology, Beijing, China
2007-08 Intern at National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
Beijing, China
2012 Senior Research Intern, Canon U.S.A., Irvine, CA
2008-13 Ph.D., University of Central Florida, Orlando, FL
2013 Senior Engineer, Qualcomm, San Diego, CA



UNIVERSITY OF CENTRAL FLORIDA
CENTER FOR RESEARCH IN COMPUTER VISION

FINAL ORAL EXAMINATION

OF

YANG YANG
B.S., BEIJING UNIVERSITY OF TECHNOLOGY, 2008

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
(ELECTRICAL ENGINEERING)

Monday, July 1, 2013, 10:00 A.M.
103 Harris Corporation Engineering Center

DISSERTATION COMMITTEE

Professor Mubarak Shah, *Chairman*
Professor Niels da Vitoria Lobo
Professor Rahul Sukthankar
Professor Gita Reese Sukthakar
Professor Kenneth Stanley

OUTLINE OF GRADUATE STUDIES

Major: Electrical Engineering

Computer Vision Systems	Shah
Computer Vision	Tappen
Advanced Computer Vision	Shah
3D Computer Vision	Foroosh
Intelligent Systems	Sukthankar
Pattern Recognition	Georgiopoulos

SELECTED PUBLICATIONS AND PATENTS

“Semi-supervised Learning of Feature Hierarchies for Object Detection in a Video”, Yang Yang, Guang Shu, and Mubarak Shah, *CVPR 2013*.

“Complex Event Detection Using Data-driven Concept”, Yang Yang, and Mubarak Shah, *ECCV 2012*.

“Learning Semantic Features for Action Recognition via Diffusion Map”, Jingen Liu, Yang Yang, and Mubarak Shah, *CVIU 2012*.

“Discovering Motion Primitives for Unsupervised Grouping and One-Shot Learning of Human Actions, Gestures, and Expressions”, Yang Yang, Imran Saleemi, and Mubarak Shah, *PAMI 2012*.

“Video Scene Understanding with Multi-Scale Analysis”, Yang Yang, Jingen Liu, and Mubarak Shah, *ICCV 2009*.

“Learning Semantic Visual Vocabularies Using Diffusion Distance”, Jingen Liu, Yang Yang, and Mubarak Shah, *CVPR 2009*.

“Systems and Methods for Generating a High-level Visual Vocabulary”, Yang Yang, Bradley Denney, Juwei Lu, et. al., *13/830,247, 2800-13976-CINC. (Pending)*

“Systems and Methods for Feature Fusion”, Yang Yang, Bradley Denney, Juwei Lu, et. al., *13/829,338, 2800-13974-CINC. (Pending)*

DISSERTATION

LEARNING HIERARCHICAL REPRESENTATIONS FOR VIDEO ANALYSIS
USING DEEP LEARNING

With the exponential growth of the digital data, automated video content analysis has been drawing increasing attention of computer vision researchers. Effective modeling of objects, scenes and actions is critical for video content understanding. Recently, there has been a growing interest in deep learning, which has shown impressive results in speech and object recognition. In this dissertation, we propose several formulations and extensions of deep learning methods which learn hierarchical representations for video analysis, including complex event recognition, object detection in videos and measuring action similarity. For complex event recognition, we propose a novel unsupervised approach to discover data-driven concepts from multi-modality signals (audio, scene and motion) to describe high level semantics of videos. Our methods consists of two main components: we first learn the low-level features separately from three modalities. Then we discover the data-driven concepts based on the statistics of learned features mapped to a low dimensional space using deep belief nets (DBNs). For improving generic object detector in videos, we present a new model that learns the hierarchical object representations in a semi-supervised manner. It differs from the existing unsupervised feature learning methods in two ways: first it optimizes for both discriminative and generative properties of the features simultaneously, which gives our features better discriminative ability; second, our learned features are more compact, while the unsupervised feature learning methods usually learn a redundant set of over-complete features. For measuring action similarity, we describe a novel approach that learns the features and metrics directly from the data. We propose a generative plus discriminative learning method based on gated auto encoders to simultaneously learn the features and their associated metrics. Extensive experiments with quantitative and qualitative results on the three tasks demonstrate the superiority of our proposed models.