



Fall 2011 Colloquium
Temple University
Computer and Information Sciences
DISTINGUISHED LECTURE
Context in Image and Video Understanding
Larry S. Davis

University of Maryland, College Park

Wednesday, 11/2, 11am, TECH Center 111

Abstract: In spite of the significant effort that has been devoted to the core problems of object and action recognition in images and videos, the recognition performance of state of the art algorithms is well below what would be required for any successful deployment in robotic applications. Additionally, there are challenging combinatorial problems associated with constructing globally “optimal” descriptions of images and videos in terms of potentially very large collections of object and action models. The constraints that are utilized in these optimization procedures are loosely referred to as “context.” So, for example, vehicles are generally supported by the ground, so that an estimate of ground plane location parameters in an image constrains positions and apparent sizes of vehicles.

Another source of context are the everyday spatial and temporal relationships between objects and actions; so, for example, keyboards are typically “on” tables and not “on” cats. The first part of the talk will discuss how visually grounded models of object appearance and relations between objects can be simultaneously learned from weakly labeled images (images which are linguistically but not spatially annotated – i.e., we are told there is a car in the image, but not where the car is located). Next, I will discuss how these models can be more efficiently learned using active learning methods. Once these models are acquired, one approach to inferring what objects appear in a new image is to segment the image into pieces, construct a graph based on the regions in the segmentation and the relationships modeled, and then apply belief propagation to the graph. However, this typically results in a very dense graph with many “noisy” edges, leading to inefficient and inaccurate inference. I will briefly describe a learning approach that can construct smaller and more informative graphs for inference. Finally, I will relax the (unreasonable) assumption that one can segment an image into regions that correspond to objects, and describe an approach that can simultaneously construct instances of objects out of collections of connected segments that look like objects, while also softly enforcing contextual constraints

Bio: **Larry S. Davis** received his B.A. from Colgate University in 1970 and his M. S. and Ph. D. in Computer Science from the University of Maryland in 1974 and 1976 respectively. From 1977-1981 he was an Assistant Professor in the Department of Computer Science at the University of Texas, Austin. He returned to the University of Maryland as an Associate Professor in 1981. From 1985-1994 he was the Director of the University of Maryland Institute for Advanced Computer Studies. He is currently a Professor in the Institute and the Computer Science Department, as well as Chair of the Computer Science Department. He was named a Fellow of the IEEE in 1997.

Prof. Davis is known for his research in computer vision and high performance computing. He has published over 100 papers in journals and 200 conference papers and has supervised over 25 Ph. D. students. During the past ten years his research has focused on visual surveillance and general video analysis. He and his students have developed foundational methods for detection and tracking of people and vehicles in video, representation and recognition of human movements and activities, and mixed AI/signal processing models for event modeling and recognition. His research has been supported by IARPA (VACE program), DARPA (VIRAT), ONR (Face recognition MURI and Compressive sensing MURI, basic research on visual surveillance), ARL (Autonomous robots) and NSF. He was a member of DARPA’s ISAT from 2007-2009 and ran an ISAT study on Persistent Visual Surveillance with Prof. Trevor Darrel of M.I.T. He is an Associate Editor of the International Journal of Computer Vision and an area editor for Computer Models for Image Processing: Image Understanding. He has served as program or general chair for most of the field’s major conferences and workshops, including the 5th International Conference on Computer Vision, the 2004 Computer Vision and Pattern Recognition Conference, the 11th International Conference on Computer Vision held in 2006, and the 2010 Computer Vision and Pattern Recognition Conference.