Chester F. Carlson
CENTER for IMAGING SCIENCE
SEMINAR SERIES

R·I·T | College of Science

www.cis.rit.edu/seminar
for schedule, abstracts,
biographies, and video archives

Computer Vision with a Third Eye

Dr. Jiebo Luo

With the increasing availability of satellite imagery of the earth, we can now perform computer vision tasks as if we had a third eye up above. First, we will present an interesting framework for event recognition. Semantic event recognition based only unconstrained still images available on the Internet or in personal repositories is a challenging problem. With GPS information, we obtain satellite images corresponding to picture locations and investigate their novel use to recognize the picture-taking environment. We then combine this inference with classical vision-based event detection methods and demonstrate the synergistic fusion of the two approaches.

However, the current GPS data only identifies the camera location, leaving the viewing direction uncertain. Second, to determine the viewing direction for geotagged photos, we utilize both Google Street View and Google Earth satellite images: 1) visual matching between a user photo and any available street views in the vicinity determine the viewing direction, and 2) when only an overhead satellite view is available, near-orthogonal view matching between the user photo and satellite imagery computes the viewing direction. Third, we explore using the phone-captured image for localization as it contains more context information than the embedded sensory GPS coordinates. The proposed approach is able to provide a comprehensive set of accurate geo-context based on the current image and its associated sensory GPS location. The geo-context includes the real location of mobile user and scene, the viewing angle, and the distance between the user and scene.

Finally, we take advantage of the aforementioned techniques to build applications to enable people to enjoy ubiquitous location-based services (LBS) using their phones. Specifically, we first perform joint geo-visual clustering in the cloud to generate scene clusters, with each scene represented by a 3D model. The 3D scene models are then indexed using a visual vocabulary tree structure. The phone-captured image is used to retrieve the relevant scene models, then aligned with the models, and further registered to the real-world map. We showcase three novel applications: 1) accurate self localization, 2) collaborative localization for rendezvous routing, and 3) routing for photographing. The evaluations through user studies indicate these applications are effective in facilitating the perfect rendezvous for mobile users.

4PM, WEDNESDAY, SEPTEMBER 19, 2012

Carlson Auditorium, Center for Imaging Science (Bldg. 76)