Visual analysis of dense crowds is particularly challenging due to large number of individuals, occlusions, clutter, and fewer pixels per person which rarely occur in ordinary surveillance scenarios. This dissertation aims to address these challenges in images and videos of extremely dense crowds containing hundreds to thousands of humans. The goal is to tackle the fundamental problems of counting, detecting and tracking people in such images and videos using visual and contextual cues that are automatically derived from the crowded scenes.

For counting in an image of extremely dense crowd, we leverage multiple sources of information to estimate the number of individuals present in the image. Our approach relies on multiple sources such as low confidence head detections, as well as texture and frequency-domain analysis to estimate counts in an image region. Furthermore, we employ a global consistency constraint on counts using Markov Random Field which caters for disparity in counts in local neighborhoods and across scales. Besides counting, we also propose to localize humans by finding repetitive patterns in the crowd image. Starting with detections from an underlying head detector, we correlate them within the image after their selection through several criteria: in a pre-defined grid, locally, or at multiple scales by finding the patches that are most representative of recurring patterns in the crowd image. Finally, the set of generated hypotheses is selected using binary integer least squares with Special Ordered Set Type 1 constraints.

Detection of complete humans in low to medium density crowds is another important problem in the analysis of crowded scenes as it is a prerequisite for many other visual tasks, such as tracking, action recognition or anomaly detection. For that, we propose to explore context in dense crowds in the form of locally-consistent scale prior which captures the similarity in scale in local neighborhoods with smooth variation over the image. Using the scale and confidence of detections obtained from an underlying human detector, we infer scale and confidence priors using Markov Random Field. In an iterative mechanism, the confidences of detections are modified to reflect consistency with the inferred priors, and the priors are updated based on the new detections. The final set of detections obtained are then reasoned for occlusion using Binary Integer Programming where overlaps and relations between parts of individuals are encoded as linear constraints.

Once human detection and localization is performed, we then use it for tracking people in dense crowds. The approach begins with the automatic identification of prominent individuals from the crowd that are easy to track. Then, we use Neighborhood Motion Concurrence to model the behavior of individuals in a dense crowd. These two aspects are then embedded in a framework which imposes hierarchy on the order in which positions of individuals are updated. The results are reported on eight sequences of medium to high density crowds and our approach performs on par with existing approaches without learning or modeling patterns of crowd flow.

Major: Computer Science

Educational Career:
Bachelor's of Computer Engineering, BS, 2007, Lahore University of Management Sciences

Committee in Charge:
Mubarak Shah, Chair, Center for Research in Computer Vision (CRCV)
George Atia, EECS, UCF
Kenneth O. Stanley, EECS, UCF
Bahaa E. A. Saleh, CREOL, UCF
Niels da Vitoria Lobo, EECS
Approved for distribution by Mubarak Shah, Committee Chair, on October 8, 2014.

The public is welcome to attend.