Multi-view People Tracking via Hierarchical Trajectory Composition

Yuanlu Xu1, Xiaobao Liu2, Yang Liu1 and Song-Chun Zhu1

1Department of Computer Science and Statistics, University of California, Los Angeles (UCLA)
2Department of Computer Science, San Diego State University (SDSU)

Overview
Hierarchical Composition Model
Experimental Results

Motivation
a) Most existing multi-view trackers exploit a particular cue, e.g., 3D localization, appearance, motion to obtain the tracking results, while different cues may dominate different periods over object trajectories, especially for complicated scenes.
b) We are interested in automatically discovering the optimal composition hierarchy for object trajectories from various cues, in order to handle a wider variety of tracking scenarios.

c) Various object scales (tracking targets sometimes either too tiny or too large).

Composition Criteria
a) Appearance Coherence: similar to bag-of-words encoding, we use a DCNN to code a person image with common people appearance templates. We fine-tune a CaffeNet (trained on ImageNet) using people image samples. The identity label is assigned by clustering. The output is regarded as an appearance descriptor.
b) Geometry Proximity: the geometric distance is measured by people’s feet positions on the 3D ground plane. We learn a kernel to handle the problem of tracklet misalignment and inaccurate feet estimation.
c) Motion Consistency: we model the motion information of a tracklet as a continuous function of its 3D ground positions w.r.t. time. We employ the b-spline function to represent the motion pattern of the trajectory, constrained by people’s relative positions.

Dataset

Features of our CAMPUS Dataset:

a) Dense foreground (around 15-25 objects, frequent conjunctions and occlusions).
b) Complex scenarios (objects conducting diverse activities, dynamic background, interactions between objects and background).
c) Various object scales (tracking targets sometimes either too tiny or too large in certain cameras).

Experimental Results

Multi-view People Tracking via Hierarchical Trajectory Composition