Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor

Wednesday, June 28, 2017 8:41 AM

- Proposes an ALFD descriptor for measuring the relative motion pattern between two detections that are separated by a time gap. The main idea can be summarized as:
1. detect FAST feature points in the two frames
2. Let f_1 and f_2 be the two frames; ALFD is a normalized sum of two unidirectional ALFDs where we will describe the one from f_1 to f_2 though the one from f_2 to f_1 can be easily formulated by symmetry.
3. For each point that lies inside the detection D_1 in f_1 and whose optical flow point is present in f_2, i.e. it was not rejected by the OF criterion, create a relative position histogram by dividing up the detection bounding box into a grid of cells and locating the cell within which the point falls.
4. To account for points that may be outside D_2, there are two additional bins that fall outside the D_2 box too - they are annular so that the first bin covers a certain distance (4 pixels) around D while the other one includes all points that are farther away.
5. If there is a 4x4 grid, there would be 16 + (16*2) bins in all since each of the 16 inner bins corresponding to D_1, there are 16+2 bins corresponding to D_2.
6. Before the binning can be started, we first need to compute the relative position of the point wrt D_1 and D_2 referred to as r_1 and r_2 → $r_1 = (P_1 - D_1)/D_1(w,h)$ where the metric is computed for both x and y directions using the width and height to normalize repetitively.
7. Once r_1 and r_2 have been computed, where $r_1 \in [0,1]$ while r_2 can be > 1 since it is not constrained to lie inside D_2, we divide up the interval into the grid and add on to the histogram based on where the values lie.
8. It is not made clear in the paper if the histogram update is purely an integral process or there is also a possibility of spreading out. So in the absence of further info we might well assume that it is an integral process so that ALFD (unidirectional) is an integral descriptor
9. Once the ALFD ρ has been computed, the affinity measure itself is given as $a_A = w_{ot} \rho$ where w_{ot} is a matrix of learned weights that is obtained during training.
10. The optimization process itself is posited as an inference problem on CRF or conditional random fields which are a type of undirected graphs.
11. A lot of heuristics have been used in the inference that are not worth mentioning. The energy function E basically has 2 terms - one for single target consistency and one for mutual exclusion so that one detection is associated with only one target.

12. Since the space of all hypotheses over which to optimize is [illegible] any practical optimization to work, a set of viable hypotheses is first generated using greedy tracklets where the detections with maximum d_p are successively added to each tracklet → the maximum d_p is taken between all detections not already in the tracklet and all that are.

13. There is also a gating strategy employed to avoid adding an excessively large number of hypotheses → predictions are generated for each target using a least squares predictor with polynomials and only these tracklets are preserved that have at least one detection with IoU exceeding some threshold.