

Machine Learning in Crowdsourcing for Video Annotation

Min Hyung Lee

Advised by Prof. Dit-Yan Yeung

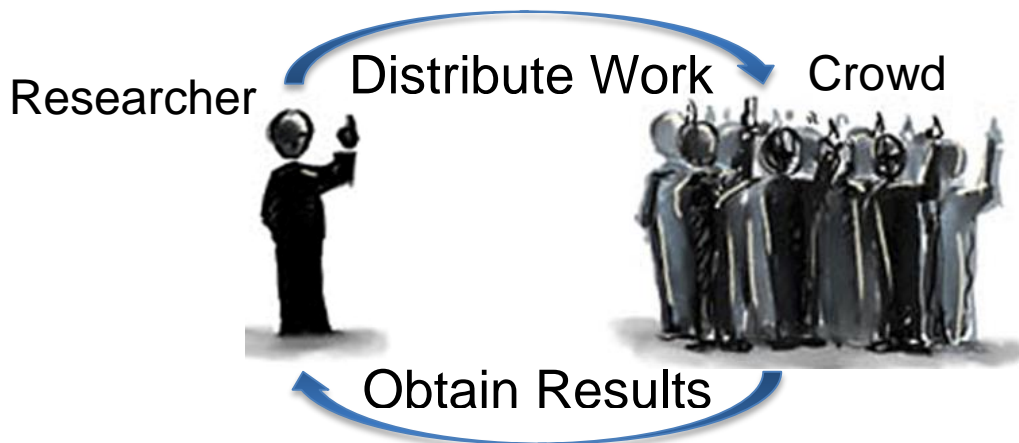
Introduction



Object trackers find the location of objects in a video

- Need labeled training data
- Labeling videos requires much cost

Crowdsourcing Platforms – Many labeled data, low cost



BUT the quality of the workers' labels is unreliable

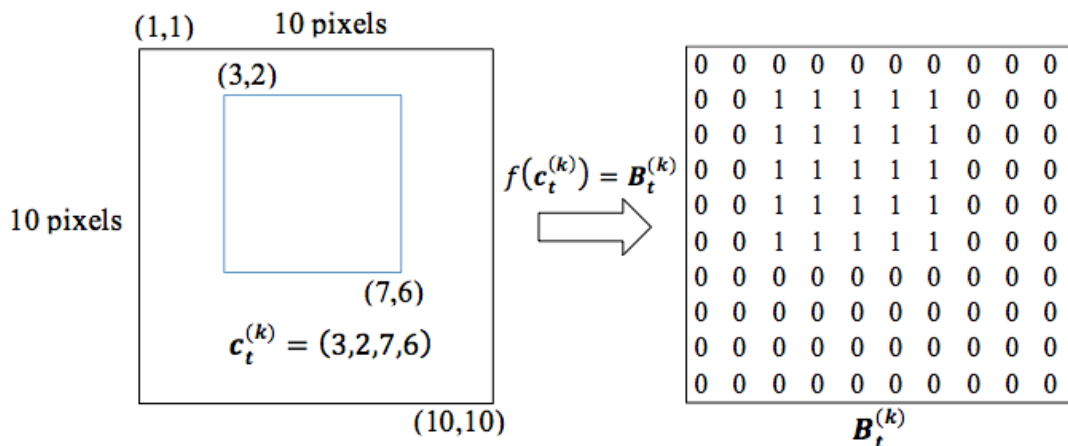
- Malicious, Inexperienced workers impair quality of labels

Objective

Use machine learning methodology to correctly aggregate the video annotation results (bounding box) from multiple workers

- The reliability of workers and ground truth label computed simultaneously

Non-Probabilistic Method

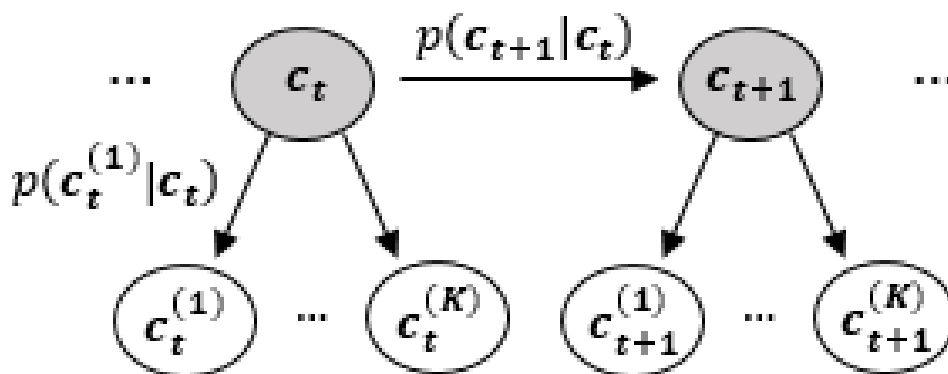


Assumption: Results closer to average are more accurate

Process

- Compute average of binary matrix
- Set weights based on distance to average
- Re-compute average

Probabilistic Method



Model using continuous-state hidden Markov model

- Probabilistically model the temporal dependency

Expectation-Maximization algorithm learns the unknown probability measures and Viterbi algorithm finds the most probable sequence of ground truth labels

Experiment Dataset

Tracking result of 9 trackers on 10 video sequences

- Trackers simulate workers in crowdsourcing setting
- Trackers are inaccurate in the latter part of video



Experiment Result

Object moving frequently and randomly

- Non-probabilistic algorithm outperform the others

Object moving slowly

- Probabilistic algorithm outperform the others

Majority of trackers perform reasonably

- Probabilistic algorithm outperforms the best tracker

