A COLLABORATIVE ALGORITHMIC FRAMEWORK TO TRACK OBJECTS AND EVENTS

Somrita Chattopadhyay, Constantine J Roros, Avinash C Kak

Purdue University

OBJECTIVE

- Track bins and their contents at airport security checkpoints.
- The pose of each bin must be tracked with precision in order to minimize the errors associated with the detection of the various items that the passengers may place in the bins and/or take out of them.

CHALLENGES

- A noisy, cluttered environment due to occlusion from passengers leaning on bins, overflowing items, etc.
- Jittery motion of the bins on the conveyor belt.
- As a bin moves away from the camera's optic axis, its appearance changes non-linearly. Thus, a traditional 2D tracker does not work.

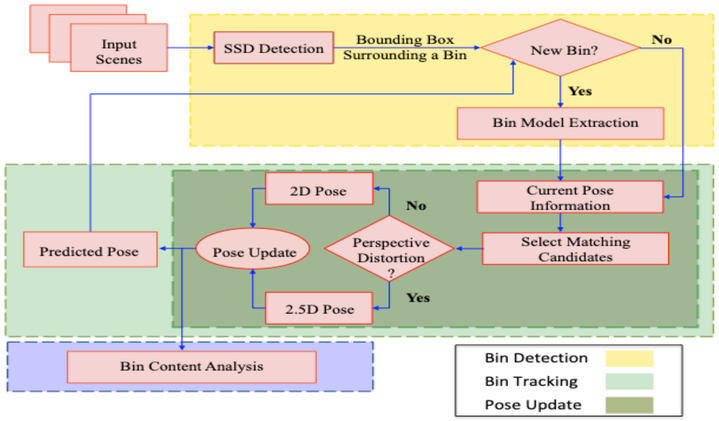
PROPOSED ALGORITHM

- **Bin Detection and Modeling**
  - Generate a bounding box around each bin using Single Shot Detector.
  - Extract straight line features inside the bounding box.
  - Fit the best rectangle to the bin boundary.
- **Tracking Initiation**
  - Tracking starts with the rectangle in the current frame as the 'initial model', if it is not being tracked already.
- **EKF based Tracking**
  - Pose Estimation
    - Find matching candidates (if any) corresponding to the projected model features.
    - A set of hypotheses is generated where each element in the set corresponds to the matching pairs of image scene and projected model features (if any).
    - A hypothesis is selected with a probability proportional to the matching pairs that are 'closer'.
    - The hypothesis is verified based on Nil-map test and the pose is updated only if the hypothesis is accepted.
  - Seamless Transition Between the 2D and 2.5D trackers
    - If the shape of the estimated bin changes significantly between two consecutive image frames, we use the 2.5D tracker which has additional state space parameters to handle the scale changes.
  - Pose Prediction
    - EKF predicts an expected view of the bin in the next frame.

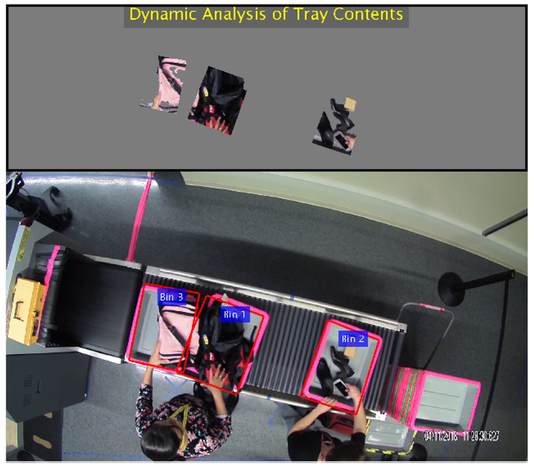
BIN CONTENT ANALYSIS

- Find Precise Locations of the Bins
  - Background subtraction using the information of an empty bin.
- Detect the Frames Where Transfer Events Occur
  - Track the bin state transitions.

OVERALL FRAMEWORK



BIN TRACKING VIDEO



Camera 9	Frame: 0573
Frame	Event
0411	Bin 1 detected
0460	Bin 2 detected
0482	Item divested in Bin 2
0500	Bin 3 detected
0515	Item divested in Bin 2
0554	Item divested in Bin 2
0555	Item divested in Bin 3
0566	Item divested in Bin 1
0573	Item divested in Bin 2

QUANTITATIVE EVALUATION

Camera #	1		2		3	
Tracker	2D	Collaborative	2.5D	Collaborative	2.5D	Collaborative
Precision % (BL)	89.1	92.1	86.7	93.5	94.5	100.0
Recall % (BL)	76.6	94.5	76.3	92.5	75.5	91.7
Precision % (TE)	85.4	96.9	86.2	93.1	100.0	100.0
Recall % (TE)	85.4	91.4	75.0	84.4	80.0	90.0

- Bin Locations (BL) : A true detection occurs if the value of an Intersection over Union (IoU) is greater than 50%, otherwise a false alarm is created.
- Transfer Events (TE): Frames within ±1 seconds are searched for an event with respect to the ground-truth.

SUMMARY AND FUTURE WORK

Our tracker estimates the bin poses accurately and produces tighter bounding boxes around the bins as opposed to other deep learning based bounding box detection. Using a more accurate pose we can track the bins and the contents within them with a higher precision. Next, we plan to incorporate a 3D EKF tracker in our collaborative framework in order to handle the scenarios where the bins undergo unpredictable 3D motion caused by the humans. We also intend to extend our collaborative framework to tracking structures other than rectangles.

ACKNOWLEDGEMENT

This research is supported by the U.S. Department of Homeland Security. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.