High-Frequency Crowd Insights for *Public Safety and Congestion Control*

Karthik Nandakumar **Sebastien Blandin** **Laura Wynter**

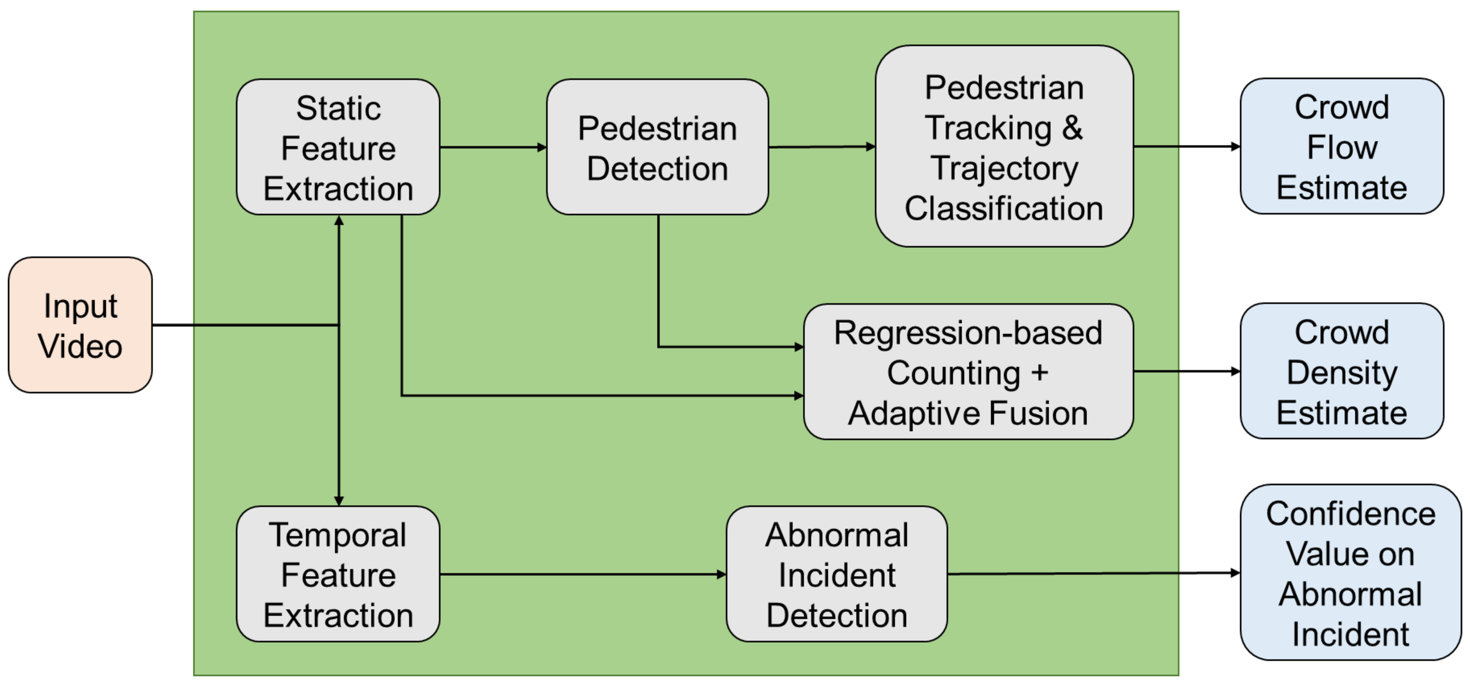
IBM

We present results from several projects aimed at real-time understanding of crowds and their behavior in urban settings. We make use of CCTV video cameras that are ubiquitous in the urban landscape. We outline the novel methods developed for our crowd insights engine and illustrate examples of its use in different contexts. Applications of the technology range from maintaining security in public spaces to quantifying the adequacy of public transport level of service.

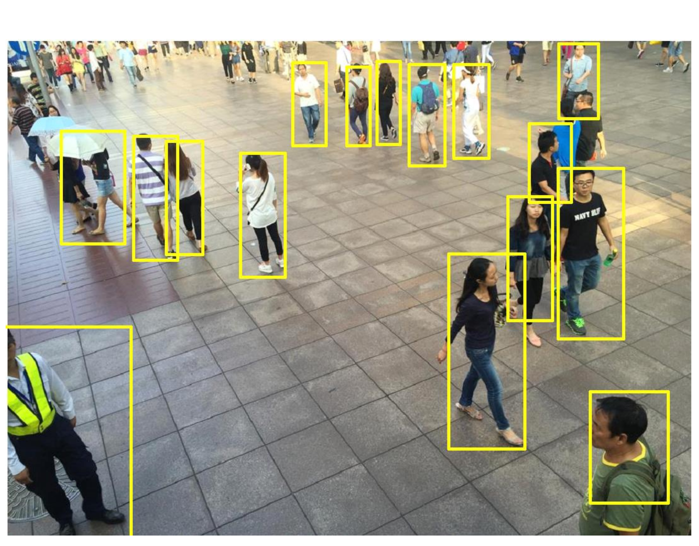
1. Introduction

- High-frequency crowd insights at events and locations such as train stations are important to support following responses:
  - provide information to individuals within & outside the crowd
  - local prescriptive enforcement via ground personnel
  - regional adjustment of resources to manage the event
- At public transport facilities, there can be dramatic variation in crowd levels (e.g., buildup in crowd level before a train arrives & rapid dispersal after the train leaves) requiring fast feedback
- Since crowding events must be precisely estimated in time, space, and magnitude as well as their nature (calm vs. angry), other types of analysis (e.g., GPS, Wi-Fi) are not sufficient because these alternatives focus only on unknown subsample

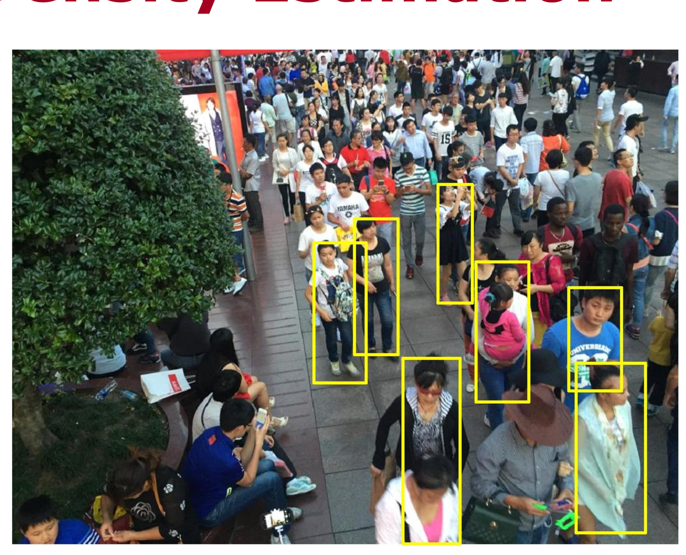
Overall Architecture of CCTV-based Crowd Insights Engine



2. Crowd Flow and Density Estimation



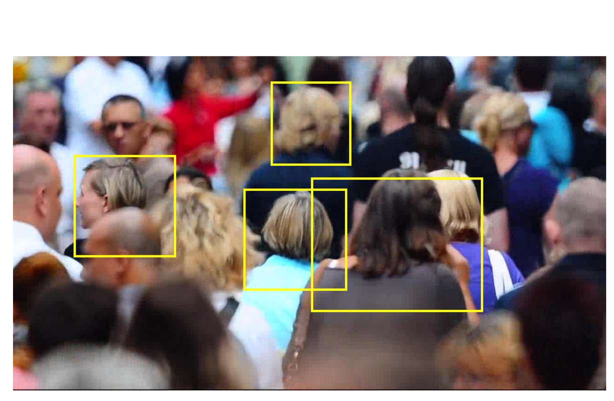
Sparse Crowd



Dense Crowd

Counting by detection approach works well for sparsely crowded scenes, but occlusion leads to severe underestimation when the crowd is dense

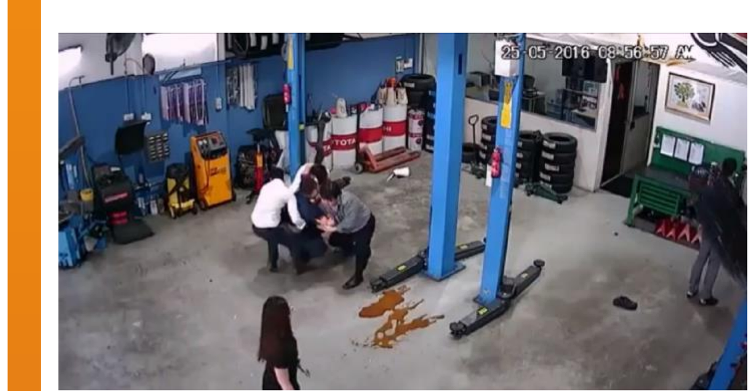
- We have developed a fusion scheme that adaptively combines counting by detection & counting by regression approaches
- Our method de-emphasizes regions with pedestrian detection (yellow boxes) & performs regression only on remaining regions



- Our method automatically assigns a higher importance to detection-based count in sparsely crowded scenes & regression-based count in densely crowded scenes
- Stringent threshold is set to minimize false detections
- People detector: Boosted decision tree classifier is applied to simple features such as HoG, gradient magnitude & texture
- Regression approach: Support vector regression is applied to 4096-d feature vector from last layer of AlexNet
- Flow estimator: Tracking of pedestrian trajectories, followed by filtering and classification into incoming & outgoing flows

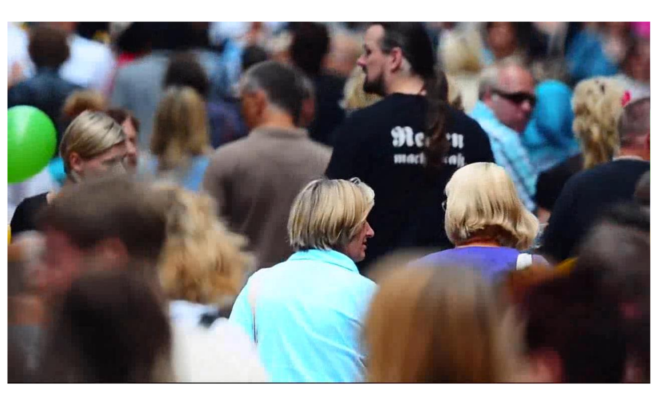
3. Crowd Incident Detection

- Crowd incidents are characterized by rapid movement of people towards each other with abrupt directional changes – tracking individuals is unreliable in crowded scenes
- Optical flow-based aggregate features characterize crowd movement patterns at coarse level

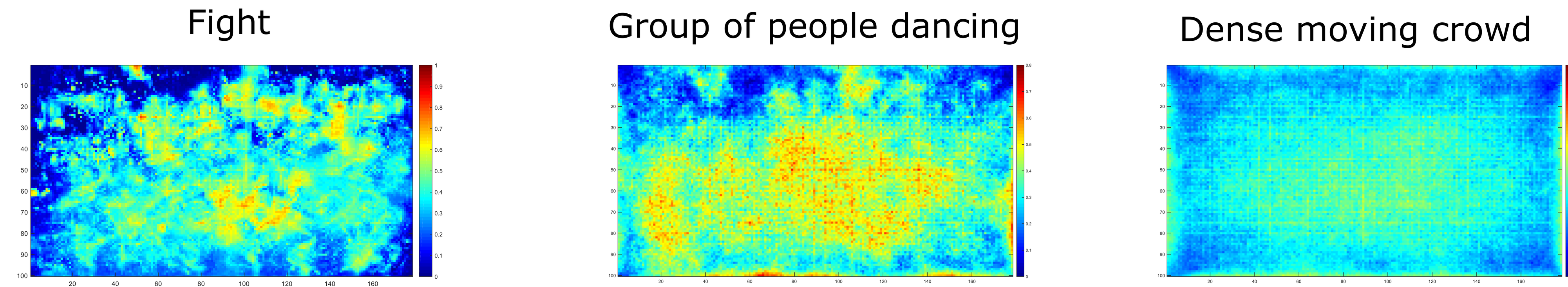


Fight

Group of people dancing



Dense moving crowd



Dark blue pixels indicate regions with low activity & red pixels denote regions with rapid motion

- Flow-based features can effectively filter out the third scenario, but they cannot distinguish between the first two scenarios because both exhibit rapid and abrupt motions
- Crowd interaction patterns [1] such as collectiveness (degree to which individuals move together) and conflict (degree to which individuals move in different directions) are useful for abnormal crowd behavior detection
- Velocities of individual tracklets (obtained by tracking automatically selected interest points for short intervals of 0-5 seconds) are used to identify crowd interaction patterns
- Two SVM classifiers are combined at score level to detect an abnormal crowd event
  - First classifier is based on histogram of block-wise aggregate flow field
  - Second classifier is based on collectiveness, conflict, and mean speed features

[1] Shao et al., "Deeply learned attributes for crowded scene understanding", CVPR 2015

4. Results

- Image annotation: Overall 420 minutes of video, from 3 urban train stations, indoor & outdoor, 4 locations (platform, gateway, lobby, escalator) during periods of heavy crowds on 2 weekdays; one day includes a train incident when the platform is fully crowded
- We annotated 63,114 ground-truth pedestrian boxes over 4,200 frames for sparse crowd scenarios, and 202,283 boxes over 6,527 frames for dense crowd scenarios; 90% of the data was used for training and 10% for testing

Ground-truth People Count	Sparse Crowd				Dense Crowd			
	[0-10]		[11-25]		[26-50]		[51-75]	
Absolute Error	Occurrences per crowd level (%)							
[0-5]	85	**100**	71	**99**	71	**92**	61	**91**
[6-10]	15	**0**	26	**1**	22	**8**	26	**9**
[11-15]	0	**0**	3	**0**	7	**0**	9	**0**
[16-20]	0	**0**	0	**0**	0	**0**	4	**0**

Confusion matrix for absolute error per crowd level. People detector is used for sparse crowd, while regression-based counting is used for dense crowd. Results from the proposed fusion model are indicated in **bold green**.

- While the error increases with the crowd level, fusion significantly improves upon the base models. These results also improve on the performance of a COTS system by 10% to 20%
- Incident detection module was tested on Violent Flows (VF) dataset [2]. With five-fold cross validation, the AUC of our system was found to be 87.9 improving on an AUC of 85 in [2]
- Since VF dataset focuses only on crowd violence, we collected a small test set of 10 videos, independently downloaded from YouTube. Among these 10 videos, 7 samples depict fights among individuals or small groups, while the remaining 3 samples depict other activities with rapid motion such as group dancing or playing games. All these 10 videos were correctly classified with an average time lag (from incident occurrence) of 10 seconds

[2] Hassner et al., "Violent flows: Real-time detection of violent crowd behavior", CVPRW 2012