

An Intent-Based Automated Traffic Light for Pedestrians

C. Ertler, H. Possegger, M. Opitz, H. Bischof

29 Nov 2018

Pedestrian Traffic Lights we're used to

- Many crossings are on-demand only
- Push-button systems
 - Pedestrians push, but **walk away**
 - Pedestrians push, but **cross on red** light
 - Many people **don't use the button** at all
 - Some push after crossing



Pedestrian Traffic Lights we're used to

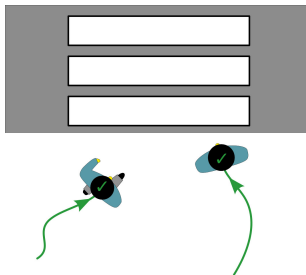
- Many crossings are on-demand only
- Push-button systems
 - Pedestrians push, but **walk away**
 - Pedestrians push, but **cross on red** light
 - Many people **don't use the button** at all
 - Some push after crossing
- Sub-optimal for traffic flow
 - Unnecessary slowdown of motorists
 - Fixed length crossing period
- Frequent replacement
 - Vandalism & impatient pedestrians
 - Substantial hardware costs per year



3 Computer Vision to the Rescue

■ Goals

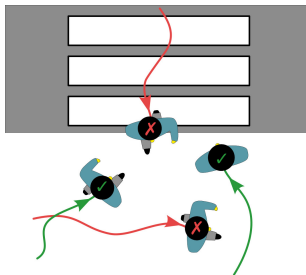
- Where is a person **heading to** and **will she cross?**
- Optimize traffic light schedule
 - Trigger based on intent
 - Reduce waiting times
 - Large groups take longer to clear safely



3 Computer Vision to the Rescue

■ Goals

- Where is a person **heading to** and **will she cross?**
- Optimize traffic light schedule
 - Trigger based on intent
 - Reduce waiting times
 - Large groups take longer to clear safely



■ Goals

- Where is a person **heading to** and **will she cross?**
- Optimize traffic light schedule
 - Trigger based on intent
 - Reduce waiting times
 - Large groups take longer to clear safely

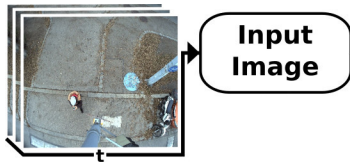


[www.agd-systems.com]

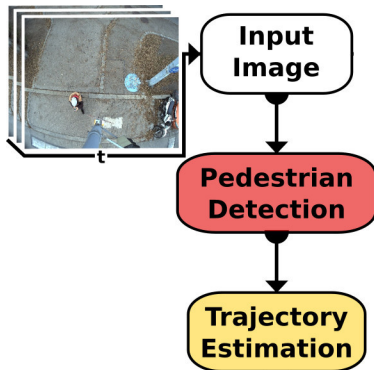
■ Commercial products

- Constrained detection region
e.g. AGD stereo sensor (2×3 m)
- Non-standard image modalities
e.g. FLIR thermal imaging sensors
- They only sense presence, but **cannot predict the intent**

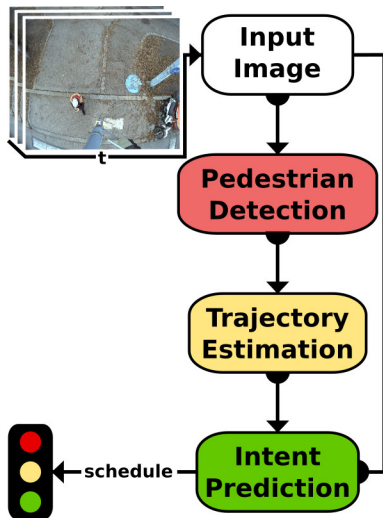
- Large field-of-view
 - Top-mounted monocular camera
 - Enables intent prediction



- Large field-of-view
 - Top-mounted monocular camera
 - Enables intent prediction
- Multi-object detection & tracking
 - Compressed detection model
 - Geometric cues for tracking

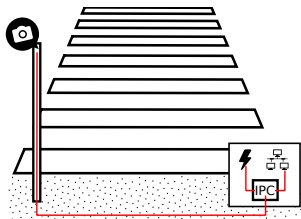


- Large field-of-view
 - Top-mounted monocular camera
 - Enables intent prediction
- Multi-object detection & tracking
 - Compressed detection model
 - Geometric cues for tracking
- Intent prediction
 - Most likely destination
 - Schedule traffic light



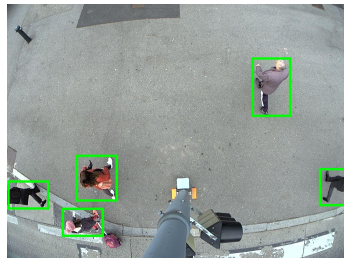
5 System Requirements

- Hardware
 - Must process locally
 - Rough environment
 - Limited space inside switch box
 - Components certified for 24/7 operation
 - Deal with frequent voltage pikes/drops
- Software
 - Robust, accurate & efficient
 - Extra system monitoring
 - Catalog of countermeasures

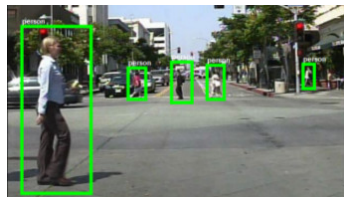


6 Multi-Object Detection & Tracking

- Accuracy/runtime trade-off
 - Single Shot MultiBox detector [1]
 - Backbone: compressed AlexNet [2,3]
 - Optimized implementation with **AVX2 instructions**
- Non-standard viewpoints
 - ImageNet pre-training
 - Application-specific fine-tuning



Our traffic light



Caltech Dataset

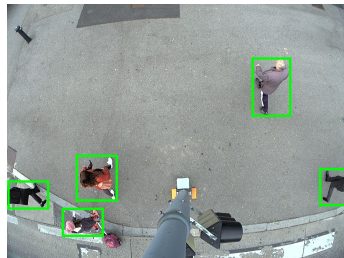
[1] Liu *et al.* *SSD: Single Shot MultiBox Detector*. ECCV'16

[2] Krizhevsky *et al.* *ImageNet Classification with Deep Conv. Neural Networks*. NIPS'12

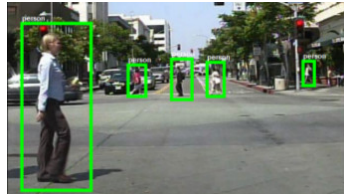
[3] Romero *et al.* *FitNets: Hints for Thin Deep Nets*. ICLR'15

6 Multi-Object Detection & Tracking

- Accuracy/runtime trade-off
 - Single Shot MultiBox detector [1]
 - Backbone: compressed AlexNet [2,3]
 - Optimized implementation with **AVX2 instructions**
- Non-standard viewpoints
 - ImageNet pre-training
 - Application-specific fine-tuning
- Tracking-by-detection
 - Kalman filters, constant velocity
 - Geometric cues & closed-world assumptions [4]



Our traffic light



Caltech Dataset

[1] Liu *et al.* *SSD: Single Shot MultiBox Detector*. ECCV'16

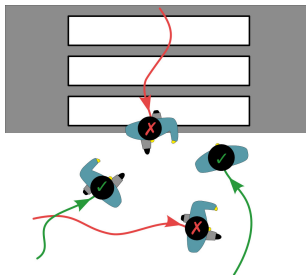
[2] Krizhevsky *et al.* *ImageNet Classification with Deep Conv. Neural Networks*. NIPS'12

[3] Romero *et al.* *FitNets: Hints for Thin Deep Nets*. ICLR'15

[4] Possegger *et al.* *Occlusion Geodesics for Online Multi-Object Tracking*. CVPR'14

7 Intent Prediction

- Where is a person headed to?
- Dynamics-based motion extrapolation
 - Exploit inertia assumption
 - Leverage instance-specific motion model
- Global motion model
 - Learn destination statistics over time
 - Accumulate priors for observed trajectory
- Combined prediction



- Robust detection on large field-of-view



- Link detections into trajectories

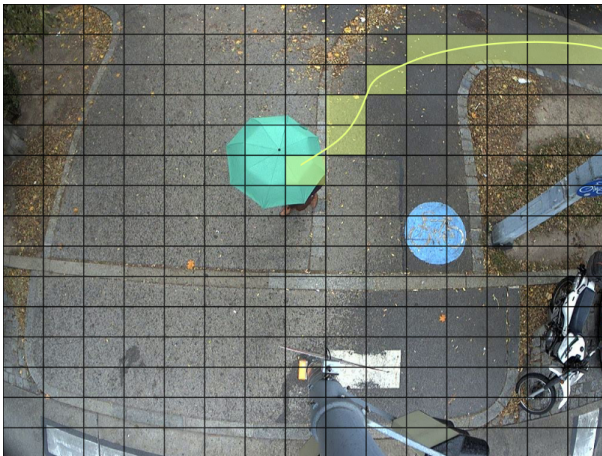


- Remove outliers, reduce jitter

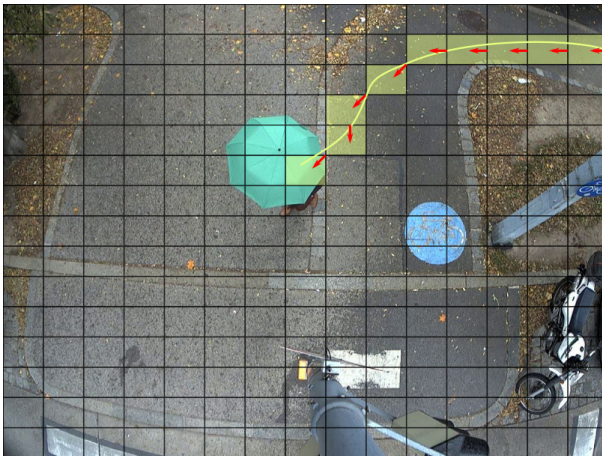


Douglas, Peucker. *Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature*. Cartographica 10(2), 1973.

- Discretize region-of-interest & rasterize trajectory



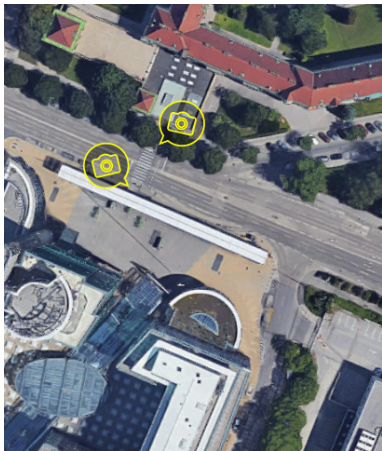
- Compute cell transitions and accumulate trajectory priors



Evaluation

Custom Dataset

- Long-term study in Vienna (Austria)
 - Both sides of a busy road
 - Over 3,000 pedestrians per day
- Recorded **1 month** of:
 - Manual push-button requests
 - (Our) intent-based triggers
 - Periodic snapshots (every 30 minutes)



Evaluation

Dataset Challenges?



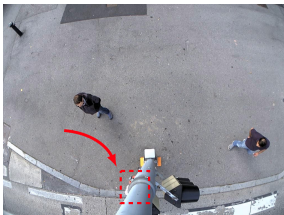
Illumination

Evaluation

Dataset Challenges?



Illumination



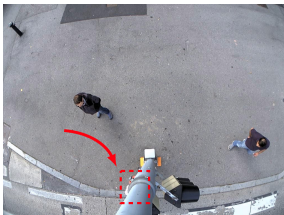
Occlusions

Evaluation

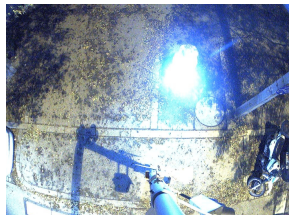
Dataset Challenges?



Illumination



Occlusions



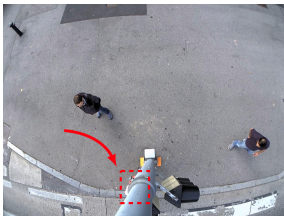
Blinding lights

Evaluation

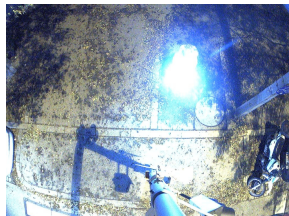
Dataset Challenges?



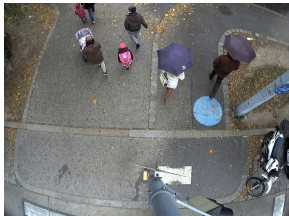
Illumination



Occlusions



Blinding lights



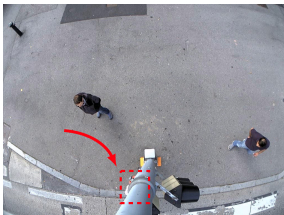
Appearance variations

Evaluation

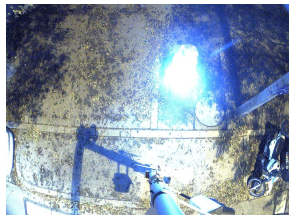
Dataset Challenges?



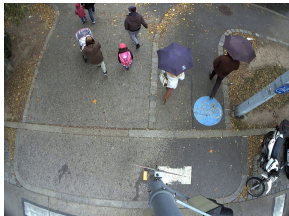
Illumination



Occlusions



Blinding lights



Appearance variations



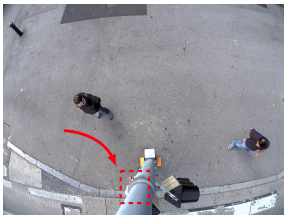
Weather

Evaluation

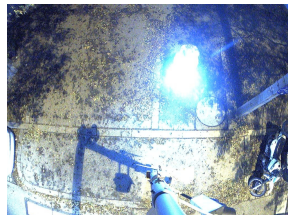
Dataset Challenges?



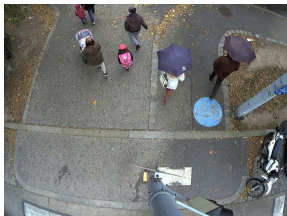
Illumination



Occlusions



Blinding lights



Appearance variations

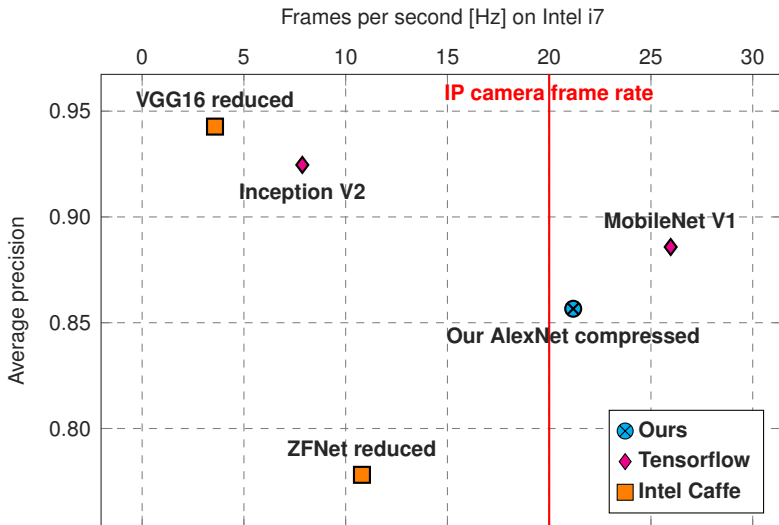


Weather



Low light at night

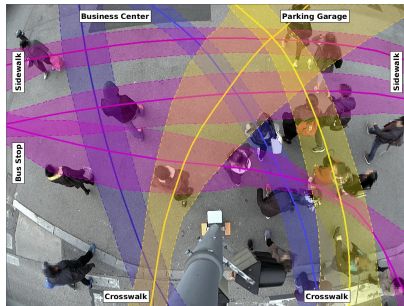
- Manually annotated over 2,000 images (60/10/30 split)



Intent Prediction

Dominant Motion Directions

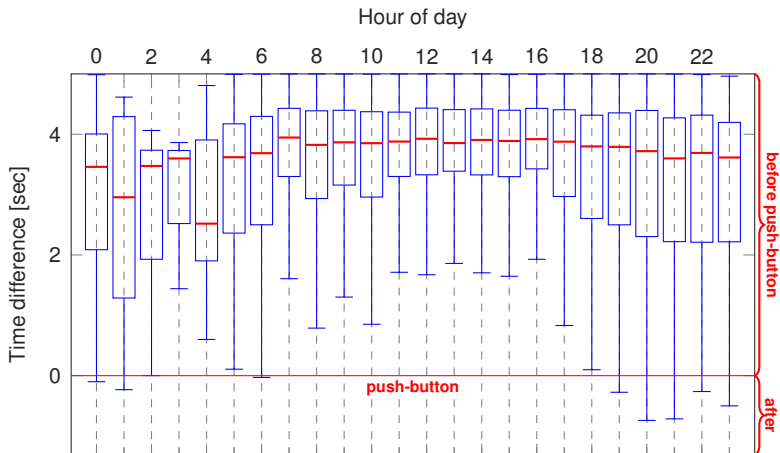
- Validated 80,000 recorded trajectories
- Shown clusters cover the majority
- Left: will a person cross or pick up her car?



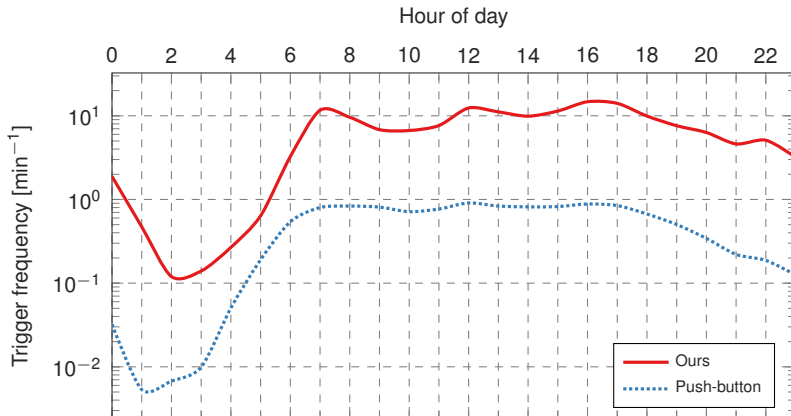
- Multi-class classification problem (predict correct destination)
- Confidence for predicted intent
- Threshold confidence to compute average accuracy

Method	Acc _{0.2}	Acc _{0.5}	Acc _{0.65}	Acc _{0.8}
EX MODEL	0.645	0.488	0.232	0.036
GM MODEL	0.977	0.969	0.924	0.492
COMBINED MODEL	0.978	0.970	0.932	0.712

- Time difference between our system and corresponding push-button triggers



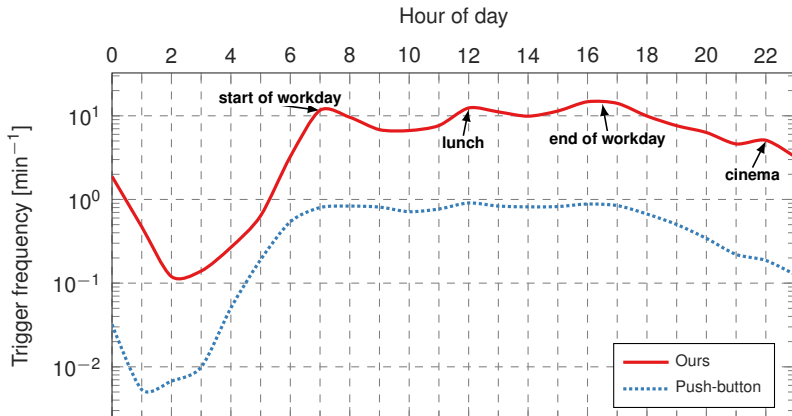
- Average trigger frequency per hour
- Our system reports 1 request per pedestrian



Intent Prediction

Reported Crossing Requests

- Average trigger frequency per hour
- Our system reports 1 request per pedestrian



- Schedule pedestrian traffic lights based on intent
 - Reliably predict crossing requests
 - Currently 3–4 sec before push-button
 - Additionally report how many people wait
 - Paved the road for further traffic flow optimizations
- Recently deployed a new system
 - Active Infrared LEDs to improve night vision
 - Processing board adjustments
 - Additional runtime improvements

Questions?



Supplemental

C. Ertler, H. Possegger, M. Opitz, H. Bischof

29 Nov 2018

Quality Measure: Intent Prediction

Last Point of Wrong Decision

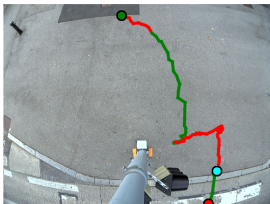
- Custom quality measure for intent prediction:
 - Last Point of Wrong Decision
 - Last Second of Wrong Decision
- Denotes the last time step a trajectory's intent prediction was wrong

Motion Model	mLPWD [%]	mLSWD [sec]
EXTRAPOLATION MODEL	0.8508	14.05
GLOBAL MOTION MODEL	0.2089	2.06
COMBINED MODEL	0.1846	1.93

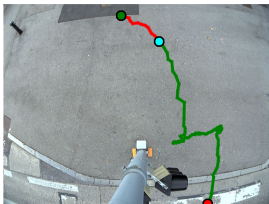
Quality Measure: Exemplary Intent Predictions

Last Point of Wrong Decision

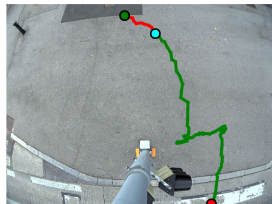
- Trajectory starts at ●, ends at ●
- Last Point of Wrong Decision ●



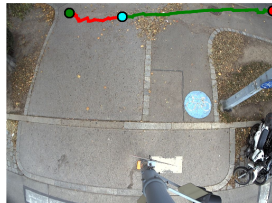
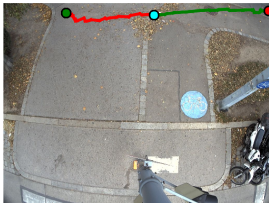
Extrapolation

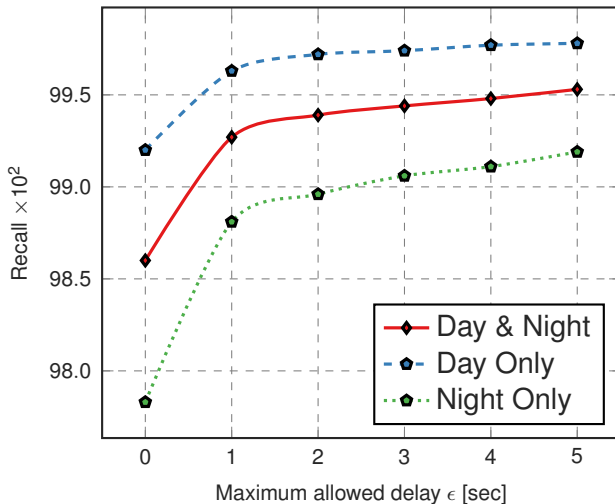


Global Motion



Combined Model





- Periodic snapshots (every 30 minutes)
- Manual annotation

TP Pedestrian wants to cross, correctly identified

TN No pedestrian wants to cross, no trigger

FP No pedestrian wants to cross, incorrectly triggered

FN Missed a pedestrian who wants to cross

	#{samples}	TP	TN	FP	FN	Recall↑	Precision↑
TRAFFIC LIGHT 1	1,578	378	1,186	11	3	0.9921	0.9717
TRAFFIC LIGHT 2	1,569	402	1,149	16	2	0.9950	0.9617
BOTH	3,147	780	2,335	27	5	0.9936	0.9665