Gradient Research & Development

Persistent Tracking for Wide Area Aerial Surveillance

Jan Prokaj and Gérard Medioni (USC)

jgradient.com/pubs/prokaj.cvpr14/
jan@jgradient.com

Problem

- 100 megapixel imagery, 1.25 frames / second, grayscale, 0.30 meters / pixel resolution
- Track vehicles as long as possible, including through stops

Approach

Detection Based Tracker
- Detections from background subtraction
- Prokaj *et al.*, CVPRW 2011

→ tracklets →

Correspondence

unmatched → initialize →

matched → update →

← tracks ←

Frame-Frame Tracker

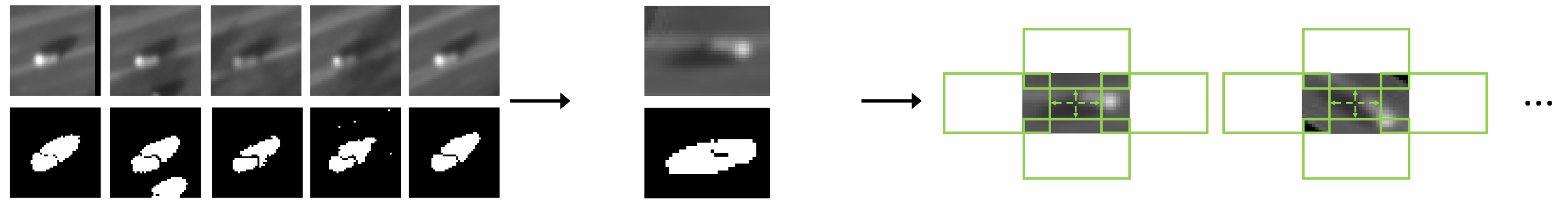
Target state regressor

Frame-Frame Tracker

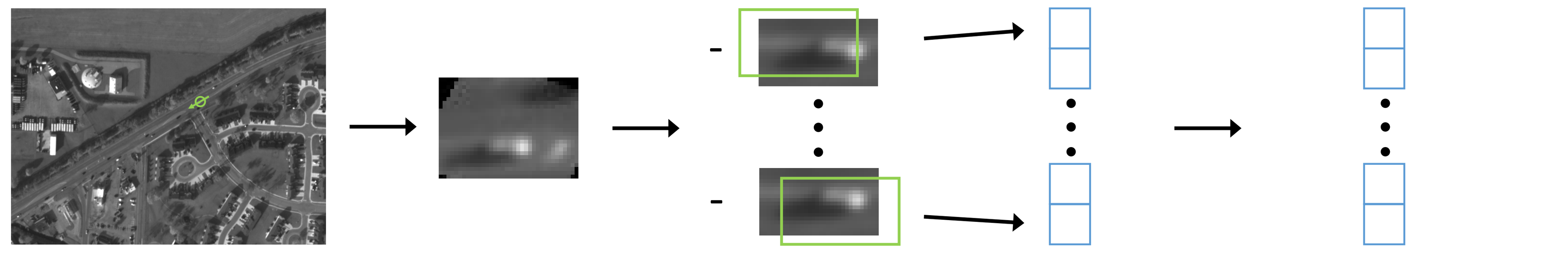
1) Sample N times from the motion model
2) Regress each sample to find displacement to true target state
3) Choose the largest cluster of target state estimates

Regression

Descriptor



Generate examples of the target from the tracklet → Transform and "average" → Generate a set of displaced versions of the template …



Given a pose sample (x,y,θ) → Transform → Subtract from the displaced template set → Normalize to a distribution

Regression

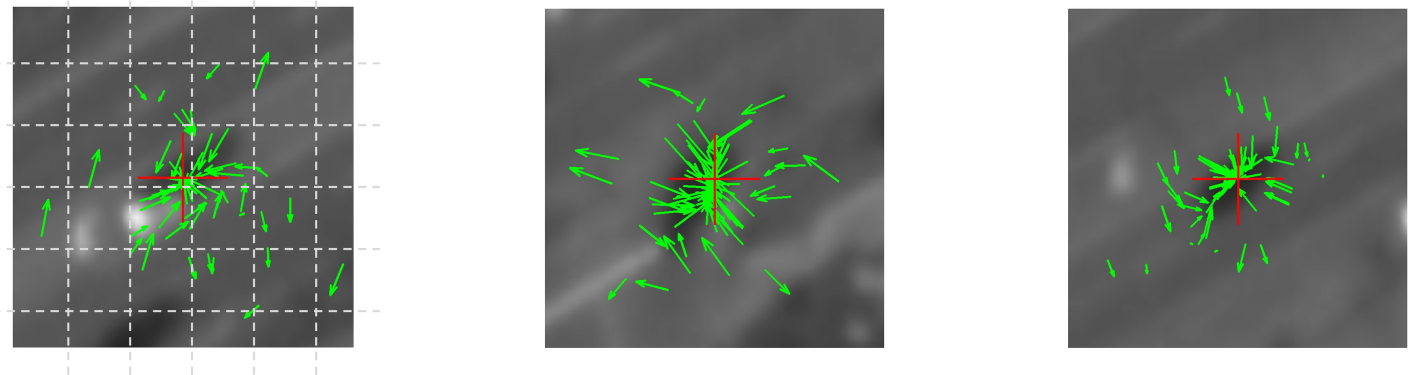
Method

- Non-parametric, Nadaraya-Watson kernel-weighted average
- Generate labelled examples from the tracklet (Δx, Δy, Δθ)
- Kullback-Leibler divergence based kernel

$$f(\mathbf{x}_0) = \frac{\sum_{i=1}^{N} k(\mathbf{x}_0, \mathbf{x}_i)\mathbf{y}_i}{\sum_{i=1}^{N} k(\mathbf{x}_0, \mathbf{x}_i)} \qquad k(\mathbf{x}_0, \mathbf{x}_i) = e^{-\lambda \mathrm{KL}(\mathbf{x}_0, \mathbf{x}_i)}$$

Validation

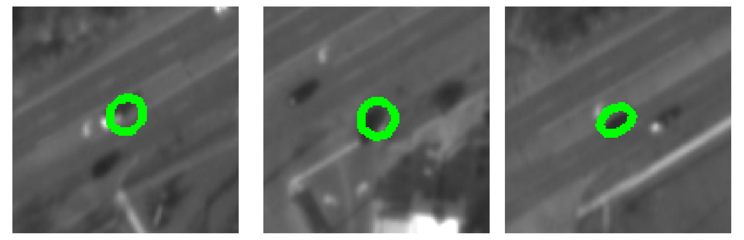
- The target must be at least partially visible for the regressor output to be valid
- Validate the output by binning the target state estimates and choosing the bin with the greatest number of samples, provided it is greater than some minimum



Motion Modelling

- Learn a vehicle motion model from a small subset of tracking ground truth
- Predict with a multi-variate RVM, combined with a velocity-dependent covariance

$$\overbrace{\underbrace{x_1\ y_1\ x_2\ y_2\ x_3\ y_3\ x_4\ y_4\ x_5\ y_5}_{\text{Features}}\ \underbrace{x_6\ y_6}_{\text{Label}}}^{\text{Vehicle trajectory}}$$

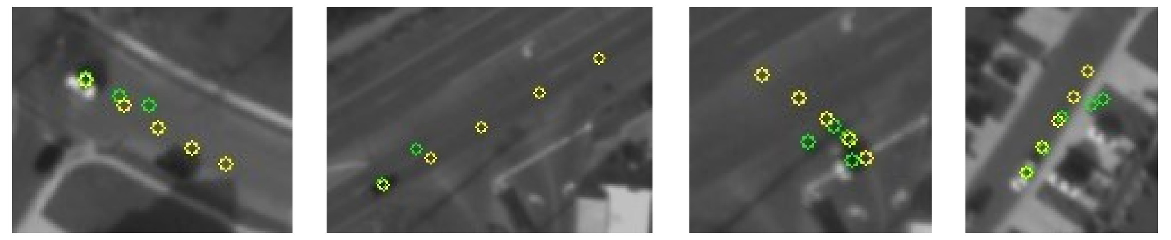


Examples of the predictive distribution

Tracker Correspondence

- Associate tracklets with tracks by considering their overlap and direction of motion
- When frame-frame tracker fails, re-initialize it with the associated tracklet

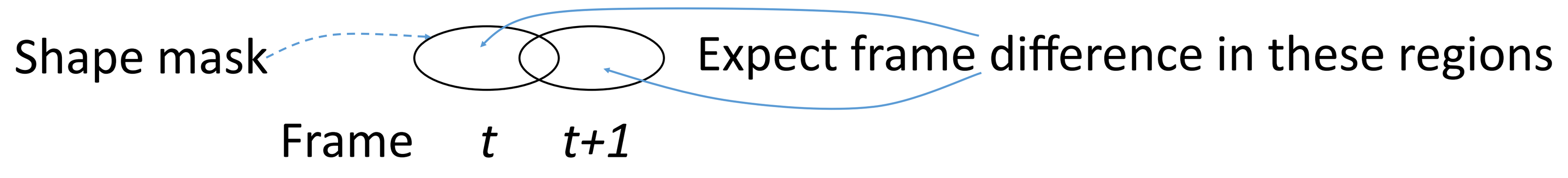
Failure Detection



Examples of diverging trajectories
- *Frame-Frame tracker* (green)
- *Detection-based tracker* (yellow)

- When the two tracker trajectories diverge, need to decide which tracker to trust
- Logistic regression classifier decides which tracker to trust using tracker confidence features (trained offline)

Motion Validation

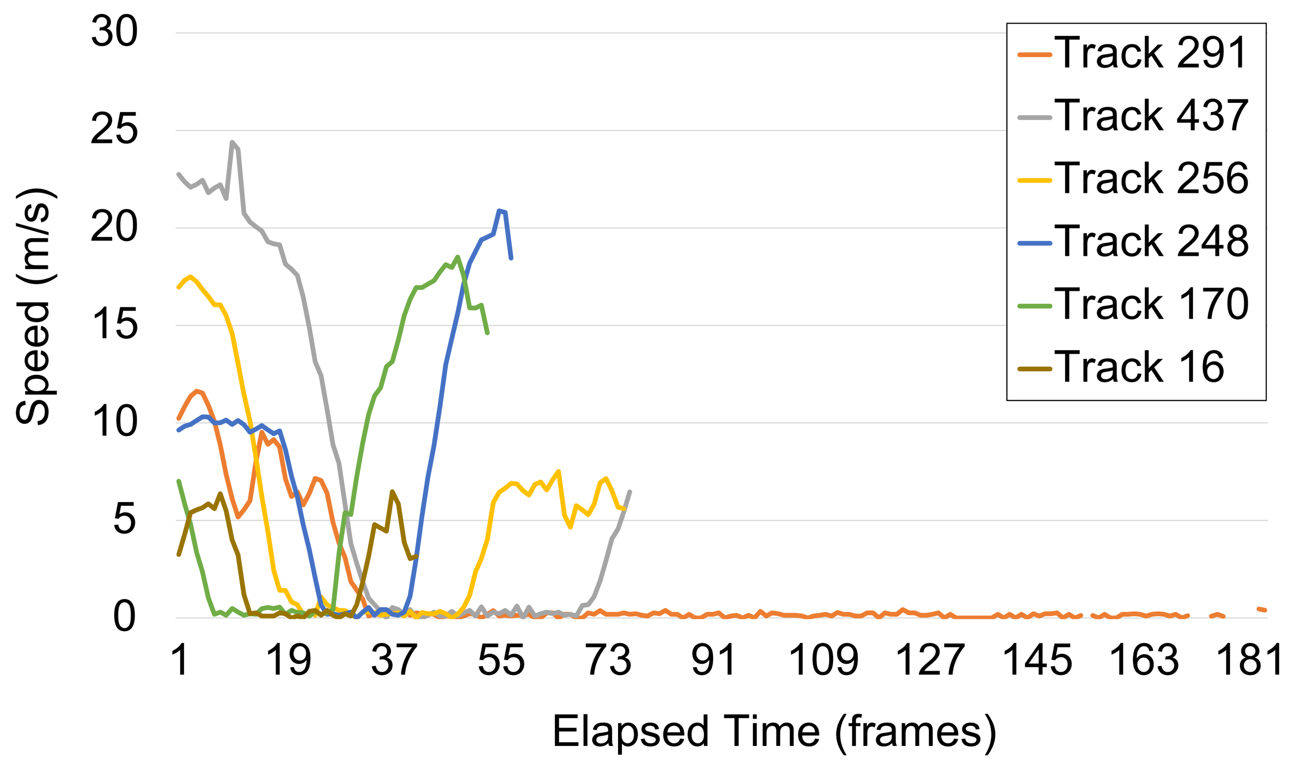


Results

- WPAFB 2009 Dataset
- 1025 frames @ 1.25 fps (≈14 minutes)
- 429 m x 429 m geo-referenced region
- 410 ground truth tracks

Generated trajectories

	Proposed	Prokaj *et al.* CVPRW 2011	Pirsiavash *et al.* CVPR 2012	Reilly *et al.* ECCV 2010	Detector Only
Recall	0.48	0.41	0.36	0.44	0.57
Precision	0.92	0.97	0.14	0.14	0.08
False Pos. / Frame	1.03	0.35	53.5	65.1	163
False Pos. / GT	0.04	0.01	2.23	2.72	6.81
MODA	0.44	0.39	-1.87	-2.27	-6.25
Swaps / Track	0.20	0.36	1.23	1.31	-
Breaks / Track	0.99	1.77	2.80	3.10	-
MOTA	0.43	0.39	-1.90	-2.30	-
Frames / Second	0.12	19.3	-	3.80	-



Examples of targets successfully tracked through a stop. The left chart shows targets' speed over time, and the image on the right corresponding trajectories.

Conclusions and Future Work

- Frame-frame tracker combined with a detection-based tracker effectively minimizes reliance on background subtraction and generates tracks of stopping targets, with little ID switching.
- Decreasing the computational complexity of the approach is the focus of our future work.