Adaptive Birth for the Generalized Labeled Multi-Bernoulli (GLMB) Filter in Satellite Videos

Camilo Aguilar*, Mathias Ortner**, Josiane Zerubia*
*Inria, Universite Côte d'Azur, Sophia Antipolis, France
**Airbus DS, Toulouse, France

Overview

01

Problem Statement

Multi Object Tracking (MOT) in Remote Sensing

Sample Remote Sensing Video

© WPAFB 2009

Dayton, OH, USA

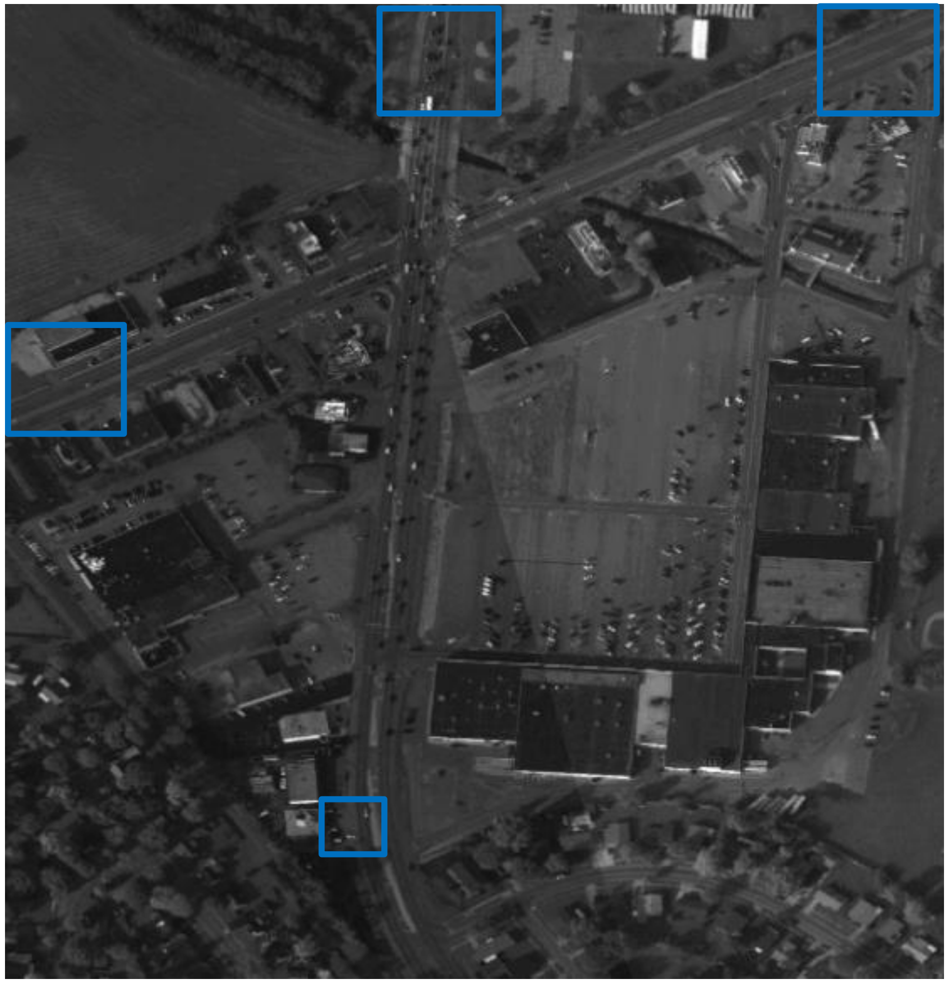
> **MOT Objective:**
- Extract target states from measurements

> **Sample Applications:**
- Urban development
- Traffic estimation
- Commerce management
- Law enforcement
- Border security

AIRBUS Ayana Inria

Challenges:

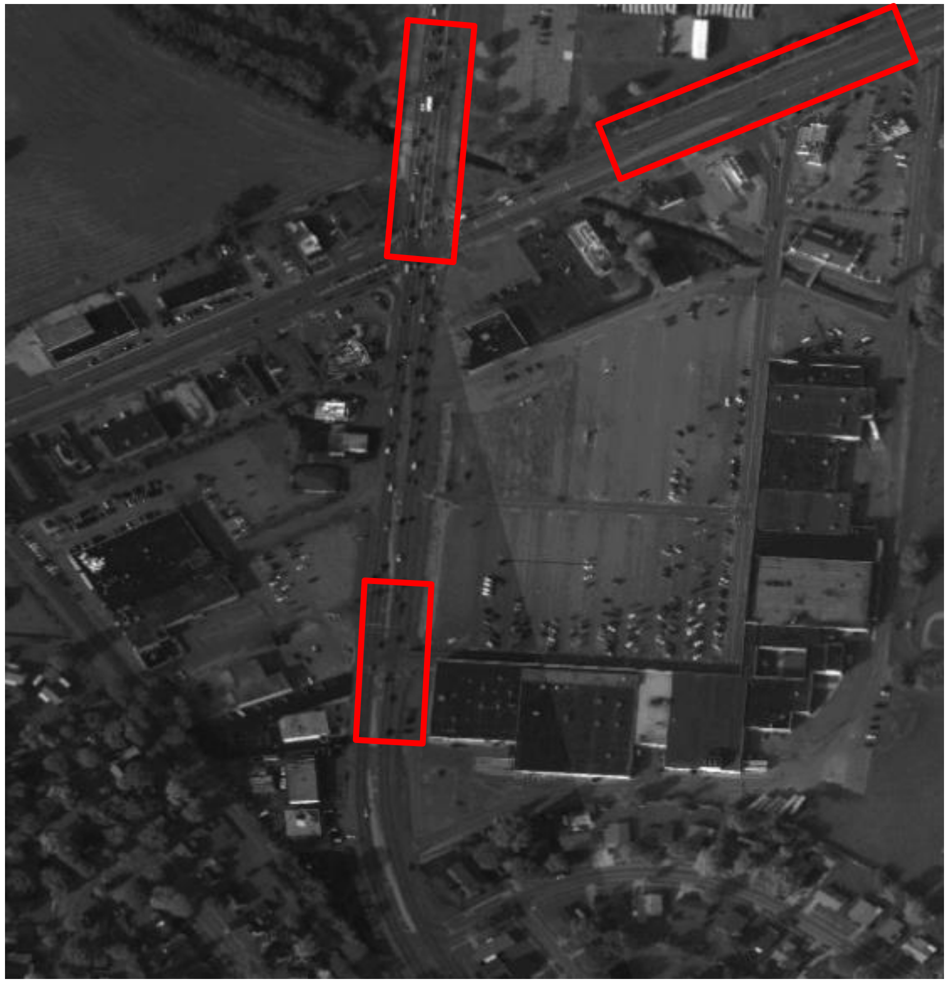
- **Varying number of targets:**
  > Targets may appear or disappear from:
    - Image borders
    - Parking lots and garages
    - Shadows and occlusions



Challenges:

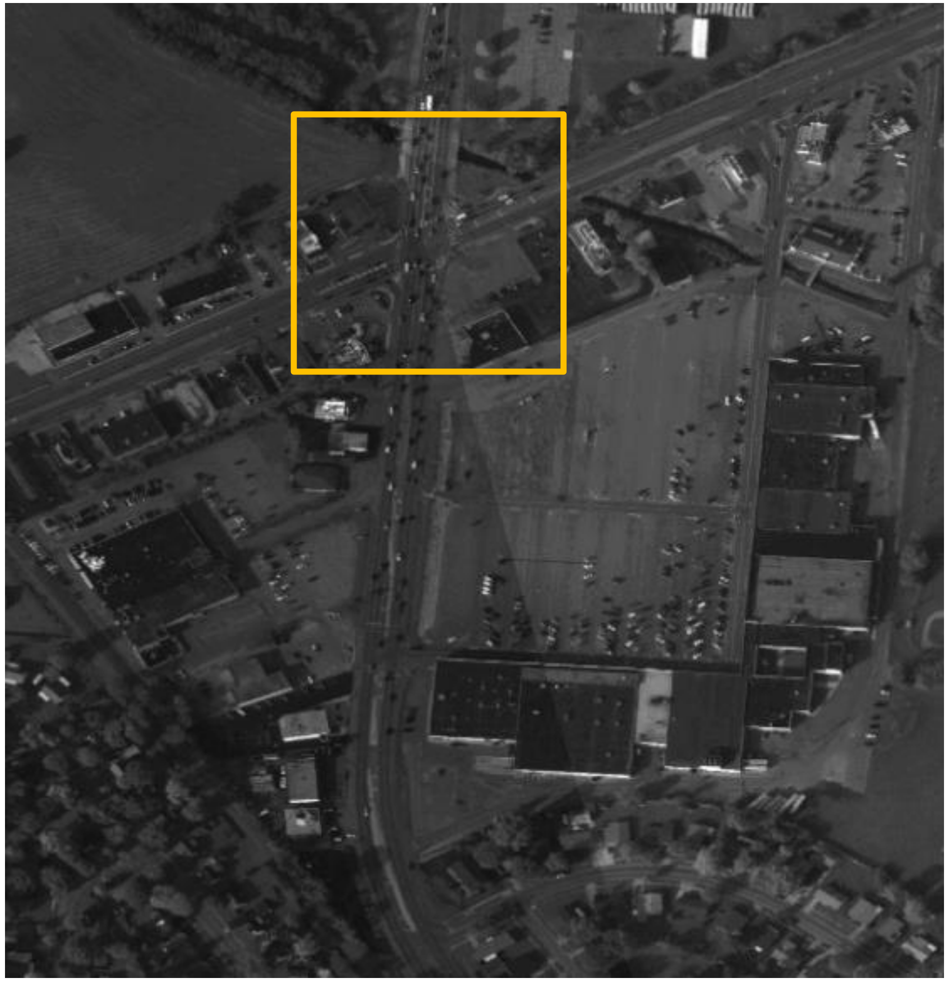
- **Varying number of targets**

- **Changing object dynamics**:
  > Objects may:
    - Move at very high speeds
    - Change directions rapidly
    - Stop at intersections



Challenges:

- **Varying number of targets**
- **Changing object dynamics**
- **Crowded scenes**:
  - > Objects can be separated a few pixels from each other



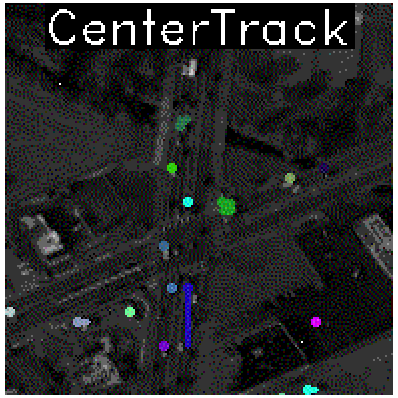
02

Popular Tracking Frameworks

Popular Tracking Frameworks

Deep Learning: CenterTrack [Zhou, 2020]

- ✓ Joint detection-tracking
- ✓ Regresses velocity vector
- x Recover tracks with ad-hoc frameworks
- x Requires 2D + time labels



Center Track [Zhou, 2020]

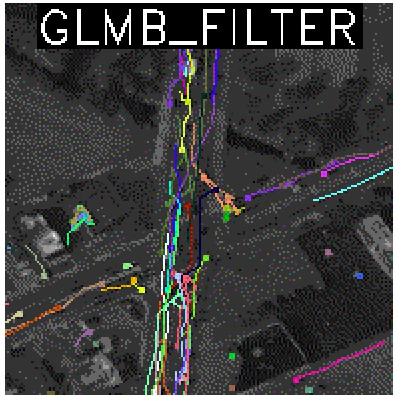
Real time:
SORT [Bewley, 2016] & DeepSORT [Wojke, 2017]

- ✓ Achieves competitive results with real-time performance
- x Relies on ad-hoc data association methods
- x DeepSORT relies on large object features

Sort [Bewly, 2016]

Random Finite Sets (RFSs):
PHD filter [Vo, 2004] and GLMB filter [Vo, 2014]

- ✓ Bayesian multi-object formulation
- ✓ Robust to clutter and misdetections
- x Requires of approximations to remain computationally feasible



GLMB [Vo, 2014]

03

The GLMB Filter

Random Finite Sets (RFS) Framework

Measurements and objects' states are modeled with RFSs [Mahler, 1994]:

Measurement Set Modeling

Single measurement: $\mathrm{z_i} = [p_i]^T$

Frame multi-measurement set: $\boldsymbol{z_k} = \{z_1^k, z_2^k, \dots, z_{M_k}^k\}$

Sequence measurement set: $Z_k = \{\boldsymbol{z_1}, \boldsymbol{z_2}, \dots, \boldsymbol{z_k}\}$

Target Set Modeling

Single object state: $\mathrm{x_j} = [p_j, v_j, w_j, h_j]^T$

Frame multi-object set: $\boldsymbol{x_k} = \{x_1^k, x_2^k, \dots, x_{N_k}^k\}$

Sequence state set: $X_k = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_k}\}$

Prediction (prior):

$$p(X_k|Z_{1:k-1}) = \int p(X_k|X_{k\,-1})\; p(X_{k-1}|Z_{1:k-1})\, \delta x_{k-1}$$

Update (posterior):

$$p(X_k|Z_{1:k}) = \frac{p(Z_{1:k}|\,X_k)p(X_k|Z_{1:k-1})}{p(Z_k|Z_{1:k-1})}$$

k: Time index
p: position
v: velocity
a: acceleration
N_k: number of objects
M_k: number of measurements

Labelled RFSs and the GLMB Filter

Labelled targets are modelled with Labelled multi-Bernoulli distributions

Target Set Modeling

Single object state: $\mathrm{x_j} = \left[p_j, v_j, w_j, h_j\right]^T$

Frame multi-object set: $\boldsymbol{x_k} = \left\{x_1^k, x_2^k, \dots, x_{N_k}^k\right\}$

Sequence state set: $X_k = \{\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_k}\}$

Prediction (prior):

$$p(X_k|Z_{1:k-1}) = \int p(X_k|X_{k-1})\ p(X_{k-1}|Z_{1:k-1})\ \delta x_{k-1}$$

Update (posterior):

$$p(X_k|Z_{1:k}) = \frac{p(Z_{1:k}|\ X_k)p(X_k|Z_{1:k-1})}{p(Z_k|Z_{1:k-1})}$$

k: Time index
p: position
v: velocity
a: acceleration
N_k: number of objects
M_k: number of measurements

AIRBUS Ayana Inria

04

Proposed Adaptive Birth

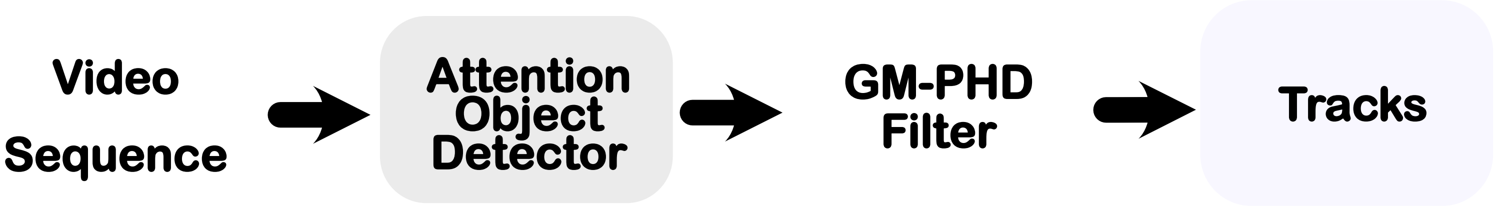
Learning a Velocity History Map h_v^k

$$
\boldsymbol{h_v^k(\hat{p}_i)} = \begin{cases} \dfrac{\boldsymbol{h_v^{k-1}(\hat{p}_i) + \hat{v}_i}}{\mathbf{2}} & \boldsymbol{if\ \left\| h_v^{k-1}(\hat{p}_i) \right\| \neq 0} \\ \boldsymbol{\hat{v}_i} & \boldsymbol{if\ \left\| h_v^{k-1}(\hat{p}_i) \right\| = 0} \end{cases}
$$

GLMB Filter with Adaptive Birth

Object Detection: Patch-based CNN

Object Tracking: GM-PHD Filter



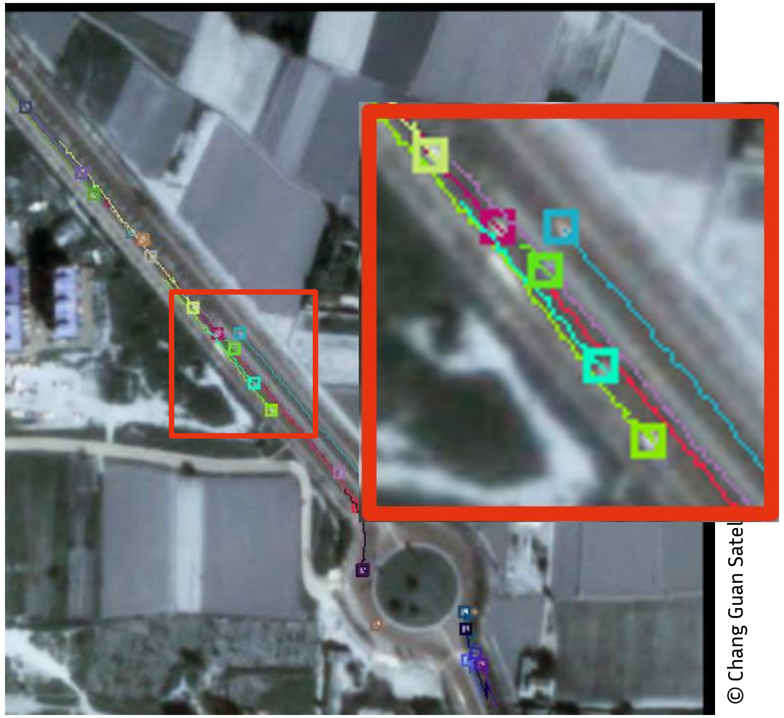
05

Results

Sample Area 1

Jilin-1 Satellite Constellation. Resolution: 1.0 m, Frequency: 20 Hz

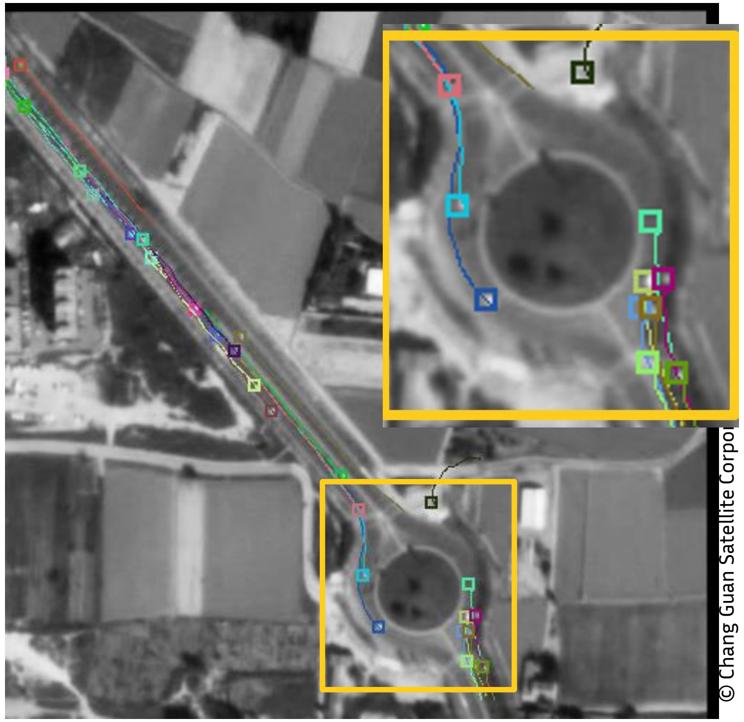
Ground Truth



Filter Output

Sample Area 1

Jilin-1 Satellite Constellation. Resolution: 1.0 m, Frequency: 20 Hz

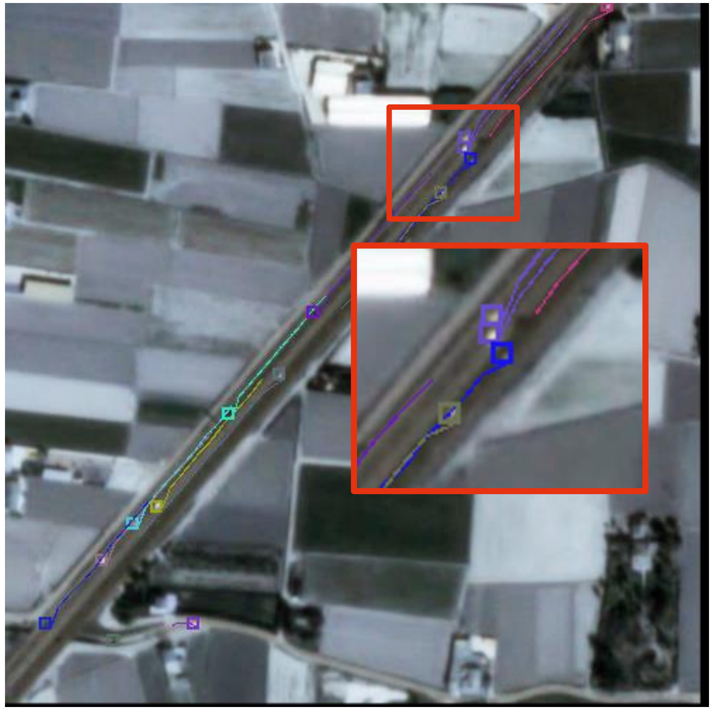


Ground Truth

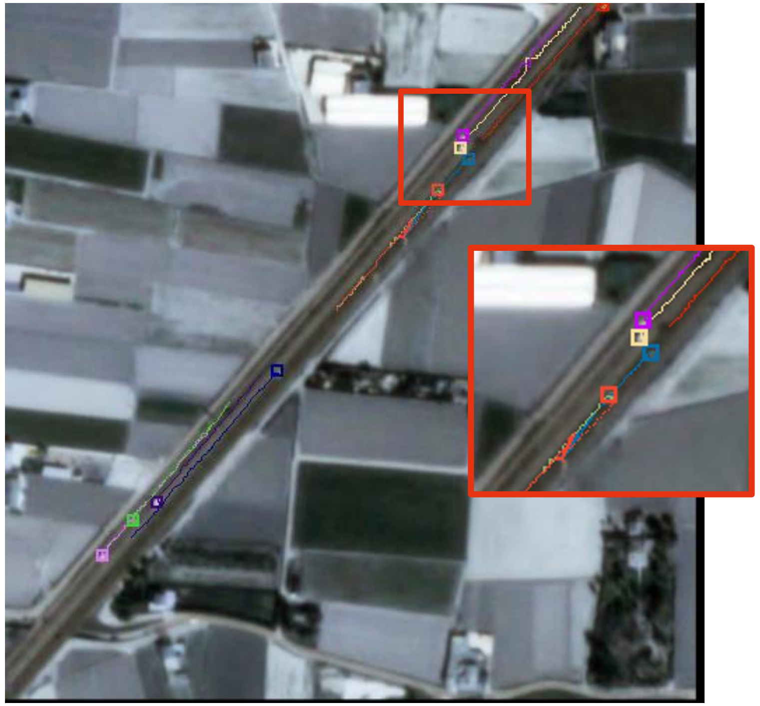
Filter Output

Sample Area 2

Jilin-1 Satellite Constellation. Resolution: 1.0 m, Frequency: 20 Hz



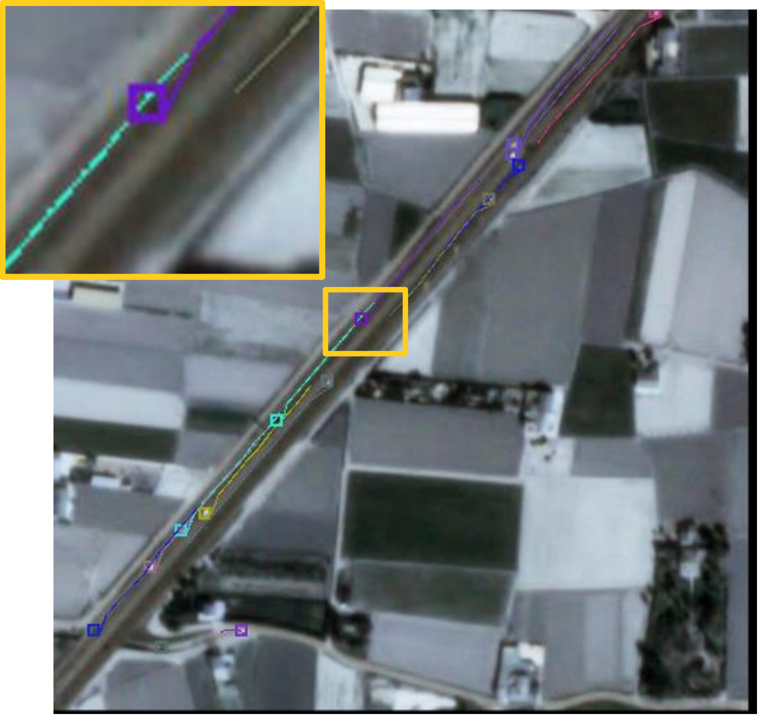
Ground Truth



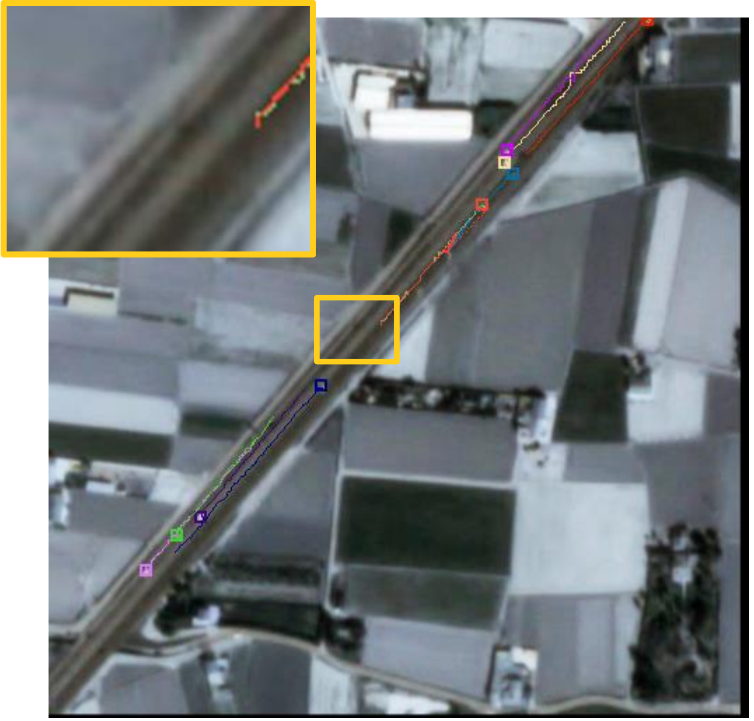
Filter Output

Sample Area 2

Jilin-1 Satellite Constellation. Resolution: 1.0 m, Frequency: 20 Hz



Ground Truth



Filter Output

ClearMOT [Bernarin, 2008] Scores

FN: False negative
FP: False positive
GT: GT objects
IDS: identity switch

d_i^k: distance to GT
N_{Frames}: number of frames
M_k: number of detected objects

MOTA: Multiple Object Tracking Accuracy

- Ranges from $(-\infty, 100]$
- Function of true positives, false positives, identity switches

$$MOTA = 1 - \frac{\sum_{k=1}^{N_{Frames}} FN_k + FP_k + IDS_k}{\sum_k GT_k}$$

MOTP: Multiple Object Tracking Precision

- Ranges from $[0.0, \infty)$
- Function of distance between detections-ground truth

$$MOTP = \frac{\sum_{k=1}^{N_{Frames}} \sum_{i=1}^{N_k} d_i^k}{\sum_k M_k}$$

Average Scores for AoIs 2, 34, 42

Method	MOTA	MOTP
Proposed	**0.871**	**0.523**
GLMB [Vo, 2017]	0.742	0.547
T-GMPHD [Aguilar, 2021]	0.464	0.554
SORT [Bewley, 2016]	0.478	0.824

05

Conclusion

Conclusion

We presented a track by detection approach to track tiny moving objects by using an attention driven CNN inference and the GM-PHD filter.

- The **patch selection mechanism** contributed to **reduce false positives** and to perform training and inference with a CNN.
- The GM-PHD filter performs a **direct multi-object approach** in comparison to common techniques of using several independent Kalman filters.
- We obtain **both better detection and tracking** scores than competing methods

References 1

[Mahler , 1994] R. Mahler. “Global integrated data fusion” , Proc. 7th Nat. Symp. on Sensor Fusion, Vol. 1, (Unclassified) Sandia National Laboratories, Albuquerque, EW Ann Arbor MI, pp. 187-199, 1994.

[Vo, 2006] B. -. Vo and W. -. Ma, "The Gaussian Mixture Probability Hypothesis Density Filter," in *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091-4104, Nov. 2006.

[Bernarin, 2008] Bernardin, Keni & Stiefelhagen, Rainer. (2008). Evaluating multiple object tracking performance: The CLEAR MOT metrics. EURASIP Journal on Image and Video Processing. 2008

[Nurhadiyatna , 2013] A. Nurhadiyatna, W. Jatmiko, B. Hardjono, A. Wibisono, I. Sina and P. Mursanto, "Background Subtraction Using Gaussian Mixture Model Enhanced by Hole Filling Algorithm (GMMHF)," *2013 IEEE International Conference on Systems, Man, and Cybernetics*, 2013.

[Droogenbroeck, 2014] Droogenbroeck, Marc & Barnich, Olivier. (2014). ViBe: A Disruptive Method for Background Subtraction.

[Ronneberger, 2015] Ronneberger, Olaf & Fischer, Philipp & Brox, Thomas. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015

[Shaoqing, 2015] Ren, Shaoqing & He, Kaiming & Girshick, Ross & Sun, Jian. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[Vo, 2016] B. Vo, B. Vo and H. G. Hoang, "An Efficient Implementation of the Generalized Labeled Multi-Bernoulli Filter," in IEEE Transactions on Signal Processing, vol. 65, no. 8, pp. 1975-1987, 15 April15, 2017, doi: 10.1109/TSP.2016.2641392.

References 2

[Feichtenhofer, 2017] C. Feichtenhofer, A. Pinz and A. Zisserman, "Detect to Track and Track to Detect," 2017 IEEE International Conference on Computer Vision (ICCV), 2017.

[Bouraffa, 2019] T. Bouraffa, L. Yan, Z. Feng, B. Xiao, Q. M. J. Wu and Y. Xia, "Context-Aware Correlation Filter Learning Toward Peak Strength for Visual Tracking," in *IEEE Transactions on Cybernetics. 2019*

[Wei, 2019] W. Ao, Y. Fu, X. Hou and F. Xu, "Needles in a Haystack: Tracking City-Scale Moving Vehicles From Continuously Moving Satellite," in *IEEE Transactions on Image Processing*, vol. 29, pp. 1944-1957, 2020]

[Beard & Vo, 2020] M. Beard, B. T. Vo and B. Vo, "A Solution for Large-Scale Multi-Object Tracking," in *IEEE Transactions on Signal Processing*, vol. 68, pp. 2754-2769, 2020, doi: 10.1109/TSP.2020.2986136.

Thank you!
Merci !
Gracias!