AerialMPTNet: Multi-Pedestrian Tracking in Aerial Imagery Using Temporal and Graphical Features

Maximilian Kraus[1,2], Seyed Majid Azimi[1,3], Emec Ercelik[2], Reza Bahmanyar[1], Peter Reinartz[1], and Alois Knoll[2]

1) Remote Sensing Technology Institute, German Aerospace Center (DLR), Wessling, Germany
2) Department of Informatics, Technical University of Munich, Munich, Germany
3) Department of Aerospace, Aeronautics and Geodesy, Technical University of Munich, Munich, Germany
Emails:{maximilian.kraus; seyedmajid.azimi; reza.bahmanyar; peter.reinartz}@dlr.de; {maximilian.kraus; seyedmajid.azimi; emec.ercelik; alois.knoll}@tum.de

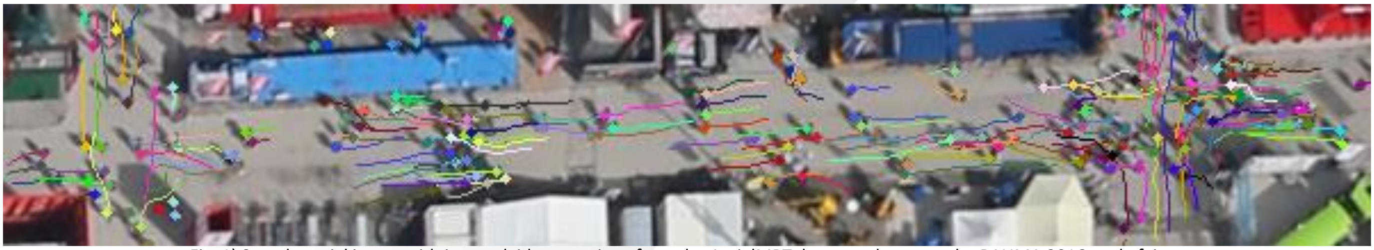

Fig. 1) Sample aerial image with its overlaid annotations from the AerialMPT dataset taken over the BAUMA 2016 trade fair.

REMOTE SENSING TECHNOLOGY INSTITUTE

1) Contributions

- We introduce a large aerial multi-pedestrian dataset, called DLR's AerialMPT.
- We propose AerialMPTNet, a CNN-based multi-pedestrian tracking network.

2) Aerial Multi-Pedestrian Tracking

Multi-pedestrian tracking in aerial imagery has varieties of applications such as in large-scale event monitoring, disaster management, predictive crowd dynamic modeling and analysis.

Aerial images can cover wide open areas allowing to analyze the behaviors of moving crowd time- and cost-efficiently.

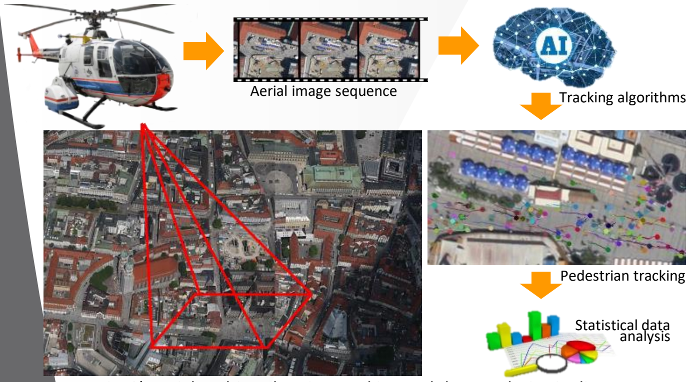

Fig. 2) Aerial multi-pedestrian tracking and data analysis pipeline.

3) Challenges and Complexities

Large pedestrian number . Tiny size of the pedestrians (e.g., 4x4 pixels) . Similar appearances of the pedestrians from above . Different image scales . Various atmospheric conditions . Extremely low frame rates (e.g., 2 fps) . Lack of annotated datasets . Limited number of task-specific aerial image tracking methods.

4) AerialMPT dataset

AerialMPT is a large multi-pedestrian tracking dataset. The images has been acquired by the German Aerospace Center's 3K camera system mounted on a helicopter flying at 600 m to 1400 m. To the best of our knowledge, it is the second dataset of its kind with the following properties:

Number of Sequences	14 (8 Train/6 Test)
Number of Frames	307
Average sequence length	21.9
Number of Annotations	44,740
Average annotation/frame	145.7
Frame rate	2 fps

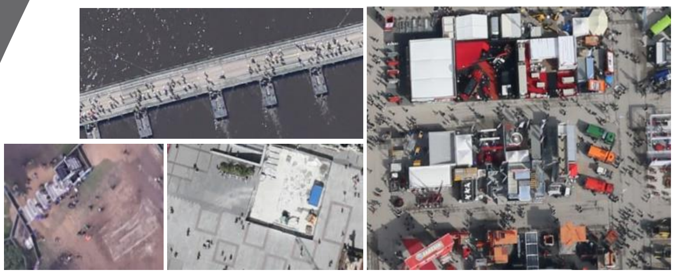

Fig. 3) Sample frames from AerialMPT.

5) AerialMPTNet

We developed AerialMPTNet for multi-pedestrian tracking in aerial images. It considers appearance features by a Siamese Neural Network (SNN) module, track histories by an Long Short-Term Memory (LSTM) module, and the neighboring objects' interconnections by a GraphCNN module (GCNN).

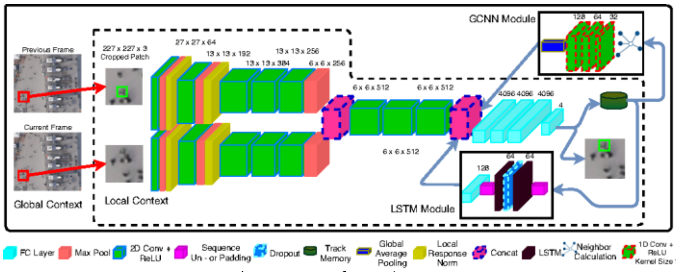

Fig. 4) Overview of AerialMPTNet.

It crops two image tiles from two consecutive frames, where in one of them object location is known and in the other one it should be predicted. The features of the two branches are then concatenated and after three convolutional layers they go through four fully connected layers to predict the coordinates of the two corners points of the bounding box around the object. The predicted object coordinates are input to the LSTM and GCNN modules, and their outputs are concatenated with the appearance features for improving the final prediction.

6) Results

We consider the widely-used MOTA metric for our evaluations, larger MOTA value is better.

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + ID_t)}{\sum_t GT_t}$$

FN (False Negative) . FP (False Positive) . ID (ID switches) . GT (Number of objects) in frame *t*

Ablation study:

Evaluating the role of different modules of the network. Each module adds value to the network performance, and all modules together achieves the best tracking performance.

	MOTA
Baseline	-37.2
Baseline + LSTM	-28.1
Baseline + GCNN	-25.4
Baseline + LSTM + GCNN	-23.4

Comparing to some other tracking methods:

AerialMPTNet outperforms traditional and other CNN-based tracking methods on the AerialMPT dataset.

	MOTA
KCF [1]	-80.5
Medianflow [2]	-77.7
CSRT [3]	-64.6
Mosse [4]	-79.3
Tracktor++ [5]	-48.8
Stacked DCFNet [6]	-41.8
SMSOT-CNN [7]	-37.2
AerialMPTNet	-23.4

7) Conclusion

Considering the specific challenges of aerial imagery allows introducing effective algorithms such as LSTM and GCNN in our work. Moreover, due to the significant importance of object tracking in aerial imagery, more efforts should be put in this domain both in the dataset generation and algorithm development.

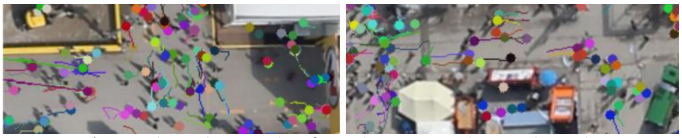

Fig. 5) Example tracking results of AerialMPTNet on the AerialMPT dataset.

References:

[1] J. F. Henriques et al., "High-speed tracking with kernelized correlation filters," IEEE TPAMI 2014.
[2] Z. Kalal et al., "Forward-backward error: Automatic detection of tracking failures," in ICPR 2010.
[3] A. Lukezic et al., "Discriminative correlation filter with channel and spatial reliability," in CVPR 2017.
[4] D. S. Bolme et al., "Visual object ttracking using adaptive correlation filters," in CVPR 2010.
[5] P. Bergmann et al., "Tracking without bells and whistles," in CVPR 2019.
[6] Q. Wang et al., "DCFNet: Discriminant correlation filters network for visual tracking,"arXiv 2017.
[7] R. Bahmanyar et al., "Multiple vehicles and people tracking in aerial imagery using stack of micro single-object-tracking CNNs," The International Archives of Photogrammetry 2019.

Deutsches Zentrum
für Luft- und Raumfahrt e.V.
in der Helmholtz-Gemeinschaft

DLR – German Aerospace Center
Remote Sensing Technology Institute
D-82234 Weßling
http://www.dlr.de/