Real-Time Target Tracking in Presence of Occlusion within Aerial Thermal Videos

Prasil Adhikari
k12049801@students.jku.at

Hamideh Ayati
k12134480@students.jku.at

Jenish Thapa
k12137169@students.jku.at

Vojtech Vlcek
k51824612@students.jku.at

Philipp Zwetkoff
k00817827@students.jku.at

Abstract

A novel unsupervised approach based on Peak signal-to-noise and Kernelized correlation Filtering.

1 Introduction

The goal of the computer vision project, titled "Real-Time Target Tracking in the Presence of Occlusion within Aerial Thermal Videos," is to develop a robust and efficient algorithm for tracking moving targets in thermal video footage captured from aerial platforms using drones. One of the main challenges in this task is the presence of occlusion, where the target is partially or fully occluded by the trees in the forest. The proposed algorithm will aim to overcome this challenge by utilizing both visual and thermal information, as well as incorporating techniques such as motion detection based on peak signal-to-noise ratio (hereinafter PSNR) and tracking by Kernelized correlation filter (hereinafter KCF) with motion estimations. The result will be a real-time target tracking system that can accurately and reliably track moving targets in the presence of occlusion.

2 Related Work

In recent years, there has been a significant amount of research in the field of target tracking in aerial thermal videos. Many of these studies have focused on developing algorithms for tracking targets in the presence of occlusion.

Some popular approaches for target detection and tracking in thermal videos use a combination of visual and thermal information, multiple hypothesis tracking, background subtraction, correlation filtering. Another notable approach that we tested for our project is deep learning-based object detection method, YOLO (You Only Look Once) in conjunction with Kalman-Filter Tracking. While YOLO and Kalman filter tracking can be effective in detecting and tracking objects in thermal videos, the supervised approach requires large amounts of labeled data, the tracking is based on the linearity assumptions and the method is also computationally expensive.

By incorporating PSNR and KCF tracker with motion estimation, our proposed unsupervised algorithm aims to overcome the limitations of YOLO and Kalman filter tracking and achieve robust and accurate target tracking in the presence of occlusion in thermal videos captured by drone.

3 Pipeline and Algorithm

3.1 Motion Detection Based on PSNR

This motion detection method was inspired by a paper written by Wei and Peng (2018) [1]. This unsupervised method was selected for the project for not needing any training dataset and thus not being limited to a certain type of objects, its robustness against noise (background motion), interpretability, wide range of adjustability and good real-time performance.

The proposed method can be decomposed into the following steps:

1. Input frames are blurred and downsampled.
2. n frames are added together to smoothen background motion and increase detection speed. This sum of n frames forms an image at time t.
3. The range of the image is scaled to a range from 0 to 255.
4. The image at time t is compared to the image at time $t-1$ to detect regions with motion. This is done in the following way:
   (a) Based on window size and stride the images at t and $t-1$ are split into patches.
   (b) For an image patch at time t and corresponding image patch at $t-1$ a PSNR is computed. The PSNR is computed as 20 times the common logarithm (with base 10) of the ration between the

maximum signal value (255) and the mean squared error between the two patches.

   (c) Only the patches below a given PSNR threshold are kept.

   (d) If window size and stride are equal (a pixel can be a member of only one patch), then pixels in the patches below the given PSNR threshold are set to 255 and the others to 0. If window size and stride are not equal (a pixel can be a member of more than one patch), then the share of memberships in the patches below the given PSNR threshold is calculated and only the pixels with the share above a given threshold are set to 255, the rest is set to 0.

5. The image with the pixels of interest is upsampled.

6. Based on the pixels of interest, regions of interest are found.

7. Regions of interest with height or width under a given minimum are ignored.

8. Regions of interest for the image at time t are proposed.

Based on the above overview of the method, certain conclusions can be drawn:

1. Motion detection and thus proposal of regions of interest starts after $2 \cdot n$ frames.

2. The proposed regions of interest change every n frames (after the first $2 \cdot n$ frames).

3.2 Tracker

The tracking algorithm is based on a paper written by Xuan et al. (2020) [2]. The authors combined a motion estimation algorithm with kernelized correlation filter to address scenarios when a target is temporarily occluded. The key parts of their proposed method are:

1. combination of a Kalman filter and motion trajectory averaging as a motion estimation method (hereinafter ME),

2. effective solution for mitigating the boundary box effect in the KCF tracker.

The first part of the ME method is a Kalman filter for estimating the position and the velocity of moving objects. This method is limited by the amount of data to converge. That is why the motion trajectory averaging (hereinafter MTA) was introduced. The MTA algorithm is based on the observation of an object's behavior in a short period of time. It assumes the position of the object at the current frame can be estimated using the speed and the position of the object in the previous frame. The amount of observations determines which method should be used first. Before the Kalman filter converges, MTA will be used as the output of ME, after convergence, the Kalman filter's result will be used.

To address the occlusion scenario, a threshold is introduced for comparison with the peak value. For the peak values lower than the threshold the object is considered occluded and the position estimation by the ME used as the position of the object. However, when the peak value of the response patch obtained by the KCF tracker is higher than the threshold, the KCF is used again to update the Kalman filter.

The tracker works hand in hand with the object detection algorithm. The proposed regions of interest by the detector are either used to initialize new tracking instances or for reconfirmation of the existing ones. An extra algorithm had to be created to combine the two approaches and enable multi-object tracking.

4 Results

This method was tested on 28 thermal videos. The proposed method demonstrated a good detection and tracking performance. The method was able to detect and track multiple objects at once. Moreover, in moderately occluded areas, approximate motion of tracked targets was well estimated. The motion detection method failed for samples with high background movement producing too many false positives. This behavior was expected due to the nature of the algorithm. Furthermore, in areas with a high degree of occlusion reliable motion estimation was not possible due to unpredictable movements of tracked targets. However, it was still able to produce rough estimates of its trajectory.

5 Limitations

The proposed method, having its advantages, also comes with certain limitations. The PSNR motion detection method is based purely on motion, which means that non-moving targets cannot be detected. Additionally, if the motion of the background is significant, it can lead to false positives or missed detections. The regions of interest proposed by this method are also not consistent in size and coverage, which can create challenges for the subsequent tracking algorithm. Furthermore, the combination of PSNR and KCF methods can lead to more false positives due to lost tracking instances, and poor tracking performance when the tracker is initialized based on an inaccurate or highly occluded region of interest. In most cases, the Kalman filter does not converge due to occluded areas and short lifespan of tracker instances. These limitations may impact the robustness and accuracy of the tracking algorithm in real-world scenarios.

6 Conclusion and Recommendations

In this project, the proposed method combines motion detection based on peak signal-to-noise ratio (PSNR) and tracking by correlation filters with motion estimations (KCF). The results of our experiments demonstrate that the proposed method can accurately and reliably track targets in the presence of occlusion, however the method also has some limitations as discussed in the above section.

In order to improve the performance of the proposed method, we recommend incorporating additional information such as depth information from stereo cameras, which can help to overcome the limitation of non-moving targets, conducting more experiments with different types of thermal videos, such as videos captured in low-light conditions, to evaluate the robustness and generalization of the proposed method, experimenting with different hyperparameters along with various preprocessing techniques. In addition, the incorporation of more sophisticated algorithms for handling occlusion and non-linear motion is certainly a future direction from this point.

Overall, the proposed method is a promising approach for target tracking in thermal videos captured by drones, but there is still room for improvement. Further research is needed to enhance the robustness and accuracy of the proposed method in real-world scenarios.

References

[1] Han Wei and Qiao Peng. 2018. A block-wise frame difference method for real-time video motion detection. *International Journal of Advanced Robotic Systems* 15, 4 (2018), 1729881418783633. https://doi.org/10.1177/1729881418783633

[2] Shiyu Xuan, Shengyang Li, Mingfei Han, Xue Wan, and Gui-Song Xia. 2020. Object Tracking in Satellite Videos by Improved Correlation Filters With Motion Estimations. *IEEE Transactions on Geoscience and Remote Sensing* 58, 2 (2020), 1074–1086. https://doi.org/10.1109/TGRS.2019.2943366