Institut für Photogrammetrie und GeoInformation

Leibniz Universität Hannover

Fakultät für Bauingenieurwesen und Geodäsie

Institut für Photogrammetrie und GeoInformation
Prof. Dr.-Ing. habil. Christian Heipke

Samer Abualhanud
Tel. +49 511 762 2488

E-Mail:
abualhanud@ipi.uni-hannover.de

9 December 2025

Image Sequence Analysis (ISA)

Lab 3: Object Tracking

Object Tracking is a key concept for many applications in the context of image sequence analysis and is subject to intensive research for decades. Tracking of objects and living beings is, for example, used as basis for traffic observation, for the generation of statistics in sports and for the analysis of the behaviour of farm animals. The goal of this exercise is that you obtain an advanced understanding of the concepts of object tracking presented in the lecture and should provide you with hands-on experience by implementing a simple tracking by detection approach.

Data provided

- Image sequence: In the context of this exercise, we will work with the image sequence that you already know from the first lab. You can download this image sequence (~250 MB) here:

  https://seafile.projekt.uni-hannover.de/f/381bebe514904006b9ad/

- Bounding box predictions: Based on the provided image sequence, pedestrians have been detected by using Faster R-CNN proposed by Ren et al. (2016) in a pre-processing step. The resulting bounding box predictions have been further processed using non-maximum suppression and exported into a pickle file. This file consists of a Python dictionary that contains one entry per frame, using the frame number as key and a list of bounding boxes predicted for this frame as value. This pickle file is named “predictions.pickle” and can be downloaded from Stud.IP.
- Visualisation function: The detected pedestrians can be visualised using the Python script “box_visualisation.py”, which can be downloaded from Stud.IP. This script takes the frames of the image sequence and the pickle file containing the predictions as inputs and creates a video file showing the predictions as red bounding boxes.

Besucheradresse:
Nienburger Straße 1
30167 Hannover
www.ipi.uni-hannover.de

Zentrale:
Tel. +49 511 762 0
www.uni-hannover.de

Tasks

1) Generate a video file showing the detected pedestrians by using the visualisation function provided. Analyse the bounding boxes shown in the video file generated. What can you observe with respect to the quality of the detections? What are common failure cases and why do they happen? Discuss and explain your findings.

2) Implement a simple tracking by detection approach. For this purpose, the following points are to be considered:

   - A detection in the current frame should be assigned to the track with the detection in the previous frame that has the smallest spatial distance in image space (w.r.t. the centre points of the bounding boxes). Make sure that every track gets assigned only a single detection form the current frame. The appearance of the pedestrians is not to be considered in this stage.
   - If a detection is not assigned to any existing track, a new track is to be initialised.
   - If a track does not get assigned a detection for a predefined number of consecutive frames τ, this track is to be ended.

   Apply your implementation on the image sequence provided, testing three different values for τ: $\tau_1=2$, $\tau_2=5$ and $\tau_3=10$. To visualise your results, take the first frame of the image sequence, showing the scene empty, and plot the computed tracks by coloured lines on this image. These lines should connect the centre points of the bounding boxes of consecutive frames that are assigned to the same track. Include the resulting three images associated to the three values of τ and the total number of individual tracks that were generated over the whole sequence in your report.

   Compare the three variants and discuss the differences visible in their results. How does the value chosen for τ influences the performance of your implementation? Which value works best and why? In which cases does the implemented approach fail? Justify your answers.

3) Implement a second variant of the tracking by detection approach described in task 2. Instead of using the spatial distance as criterion for the assignment of detections to tracks, the appearance of the pedestrians is to be used. (Note that you are also allowed to combine appearance and spatial distance.) The choice of the specific criterion is up to you. Describe the approach that you are using to assign detections to tracks and explain the reasons for your choice.

4) Test the implementation of the second variant (from task 3) following the procedure described in task 2, again using the three values for τ specified. Include the visualisation of the computed tracks and the total number of tracks over the whole sequence in your report also for this variant.

   Which differences can be observed between the variants from tasks 2 and 3? What are the reasons for these differences? What limitations remain when considering the appearance to assign detections to tracks? What potential solutions could you think of to overcome these limitations? Discuss these questions and justify your answers.

Submission details

- Groups of two students may work together on this exercise.
- The following items should be part of your submission:
  1. Your documented Python code
  2. The discussion of the questions (incl. visualisation for the results)
- The discussion is to be prepared in written form (full sentences in English) as a single PDF file containing the names and matriculation numbers of all group members.
- All items are to be submitted contained in a single zip-archive using the following naming convention: familyName1_familyName2.zip (e.g., Mueller_Meier.zip)
- This archive is to be uploaded to Stud.IP into the submission folder of exercise three.
- Deadline for submission: **06 January 2026**