1 Documentation - Multi-view Pedestrian Tracking

This document describes the dataset multi-view pedestrian tracking which was captured during the project GRK 2159: Integrität und Kollaboration in dynamischen Sensornetzen (i.c.sens). This dataset consists of 29 image sequences which capture trajectories of pedestrians from three different viewpoints with three stereo cameras. While the dataset was created mainly for the purpose of multi-views collaborative pedestrian tracking with stereo image, it can certainly be used for single view point tracking and pedestrian detection as well. A detailed description of the dataset as well as how to use it are described in the following sections.

1.1 Acquisition

The measurement campaign was carried by the team of the i.c.sens research training group, in which the author (Uyen Nguyen) took primary responsibility. Different from existing benchmarks, pedestrians are observed from various viewpoints by three stereo systems (two Pointgrey GS3-U3-23S6C-C and one Allied Vision AV MAKO G-234C) at the frequency of 10 Hz. The dataset, therefore, is called Multi-views (MuVi).

Besides capturing unknown people moving in the experimental area, a number of i.c.sens colleagues also took part in the measurement as pedestrians walking along pre-designed paths to create complicated scenarios. In this way, not only the number of people appearing in images but also the complexity of the whole scenes were partly controlled. Besides the difficulties caused by crowded scenes, the challenges of MuVi also come from illumination conditions, and image blur, which can occur when capturing image sequences in the real-world. Consequently, MuVi can be employed to test and analyse the generalization capabilities of tracking approaches and to improve their robustness and accuracy.

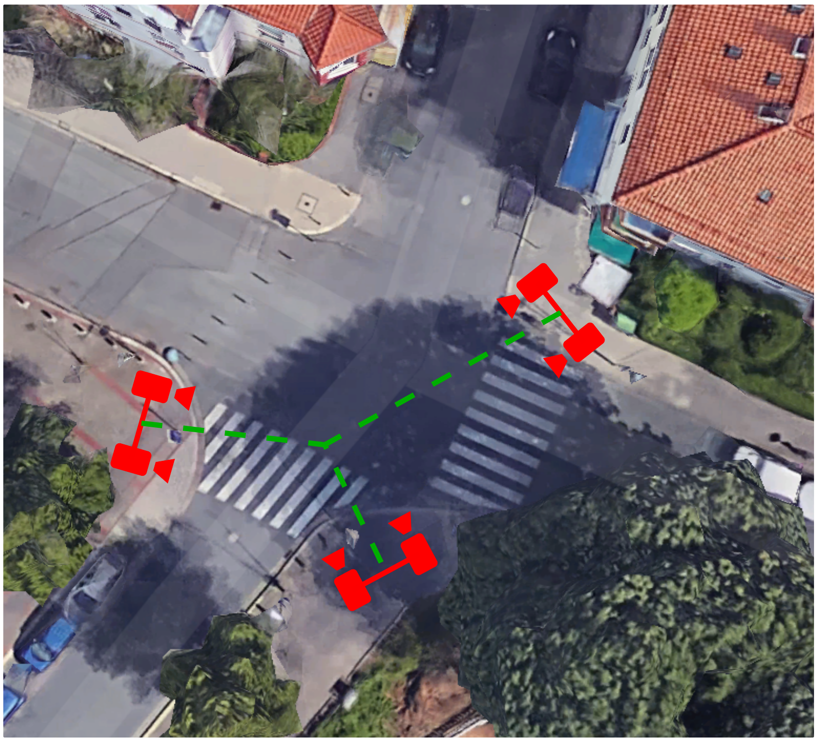

Figure 1: The overview of the scene and how the cameras were placed at an intersection.

1.2 Image data

The images from each viewpoint are stored in: *view1.zip, view2.zip, and view3.zip.* Each viewpoint has a number of different sequences that corresponds to various scenarios and movements of pedestrians in the scene. The sequences from three viewpoints have the same name if they were captured close in time and the number of images in these sequences are also equal or different not more than one image. For each sequence, there is a *left* and *right* folder which contain left and right images of the stereo system. These two images are already rectified to epipolar images. The corresponding left and right images have same name. In addition, a *timestamp.csv* file which specifies the absolute GPS time of a stereo image pair also provided. The order of the data can be describes as follow chart:

```
view*.zip
   └── (sequences)
       0000, ..., 0009
            └──(left folder -- right folder -- timestamp.csv)
                000***.png -- 000***.png -- GPS timestamp
```

1.3 Label data

The annotation of ground truth was done manually which includes 2D bounding boxes which cover the whole body of pedestrians in the left image of a stereo pair. Also an identical and unique identity is assigned for each pedestrian in the scene and across the viewpoints. The following annotations are provided in text file for each capture:

- (1) the image frame
- (2-5) a 2D bounding box of pedestrian (x-cord y-cord width height)
- (6) an ID (identical across viewpoints)
- (7) occlusion level (0, 25, 50, 75, 100 %). This evaluation may not be very consistent in the whole dataset because it depends on the people who did the labelling.
- (8) the person is a pedestrian (1) or a cyclist (0)

1.4 Calibration parameters

Since the stereo cameras were placed static during the measurement, the camera parameters are provided for each viewpoints in the *calibration.zip* file. There are three different files for each viewpoint includes: *intrinsics.txt, extrinsics.txt, absolute.txt.*

intrinsics.txt provides intrinsic camera parameters of the left and right cameras (before the rectification) which include intrinsic matrix and distortion matrix:

extrinsics.txt contains relative orientation parameters of the stereo camera. The camera parameters of the rectified images are contained in $P1$ and $P2$, respectively.

$$P1 = \begin{bmatrix} -f & 0 & x_0 & 0 \\ 0 & -f & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad , \quad P2 = \begin{bmatrix} -f & 0 & x_0 & -fb \\ 0 & -f & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

absolute.txt provides absolute orientation parameters of the left rectified stereo camera (the right camera has the same orientation but is shifted by the baselength). A point X in the global system can be projected to the left image by

$$x_{left} = K_{left} * R' * [E| - X0] * X \ .$$

A point X in the global system can be projected to the right image by

$$x_{right} = K_{right} * (R' * [E| - X0] * X + B) \ ,$$

where $B = [b, 0, 0]$.

Rotation matrix R can be calculated from the three Euler angles rpy by:

$$R = R_y * R_p * R_r$$

$$R_y = \begin{bmatrix} cos(y) & -sin(y) & 0 \\ sin(y) & cos(y) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R_p = \begin{bmatrix} cos(p) & 0 & sin(p) \\ 0 & 1 & 0 \\ -sin(p) & 0 & cos(p) \end{bmatrix}$$

$$R_r = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(r) & -sin(r) \\ 0 & sin(r) & cos(r) \end{bmatrix}$$