Second Laboratory Work: Detection and Tracking of pedestrians in public spaces

I. Introduction

Performing the detection and tracking of pedestrians is an important research field in the computer vision and image processing communities. One of the main applications is to perform the surveillance in several and quite different environments, for instance, in airports, shopping malls, subways and in other many public spaces. Performing the two above tasks (*i.e.* detection and tracking), it is possible to collect pedestrian's trajectories, and thus to perform a statistical analysis. Such statistical study can include the estimation of the occupancy in a given environment (*i.e.*, most visited places), or the estimation of most typical trajectories or activities. Also, it is possible collect unusual activity or behaviour (*i.e.*, "strange" trajectories), that corresponds to paths that deviates most from typical trajectories. This is important to detect suspicious behaviour.

However, obtaining an algorithm capable of estimating the pedestrian locations through time is challenging. Several issues contribute to turn this problem difficult: (i) the high variability that characterizes the pedestrians that includes (i-1) the *appearance* of a pedestrian in the image that is affected by the pedestrian's pose or, (i-2) the *clothing*; we also have to take into account (ii) the *illumination* conditions that are nor constant during the time acquisition; (iii) the *background clutter* and (iv) *occlusions*. All the above issues play a n important role in making pedestrian detection a challenging problem to be solved.

The goal of this work is to develop an algorithm capable to provide the pedestrian's trajectories using conventional handcrafted features.

As a final remark, notice that **it is expected that the output of the algorithm will be enriched using the visual information as much as possible**.

II. Datasets

For this work we will use the publicly benchmark datasets. Among several datasets, we will use the PETS family dataset. The dataset is available in [Dataset]. In this link we can see that several datasets are available, including:

1) Dataset S0: includes the subsets *background*, *city center* and *regular flow*
2) Dataset S1: includes the subsets **S1.L1**, **S1.L2** and **S1.L3** *walking*
3) Dataset S2: includes the subsets **S2.L1**, **S2.L2** and **S2.L3** *walking*
4) Dataset S3: includes the subsets **S3** *multiple flows* and **S3** *event recognition*

An overview of the data organization of the above datasets is available in [Overview]. Here you can find some information regarding the dataset. Important information comprises: (i) samples (ii) name of the sequence (iii) frames per second (FPS) (iv) image resolution (v) video length (vi) tracks (vii) bounding boxes.

In [Overview] we can see that each subset may include one, or more time acquisitions, denoted as *time-h-m*. Each *time-h-m* contains several *views*, numbered as *view-0*, *view-1*, ... *view-8*, see Figure in [Overview] for an illustration.

In this work, we will use the subset **S2.L1**, the first sequence mentioned in 3), with *Time-12-34* and in the *view-0*. Fig. 1 shows some images samples belonging to this **S2.L1** sequence in the *view-0*.

Figura 1. Frame samples from the sequence **S2.L1** in the *view-0*.

III. Ground truth data to measure the performance of the algorithm

The output of the algorithm should be the detections of the pedestrians. One way to accomplish this, is to use a bounding box around each pedestrian or group of pedestrians. Figure 2 shows an example of an image sample (left) and the same image with the corresponding detections represented in bounding boxes (right).

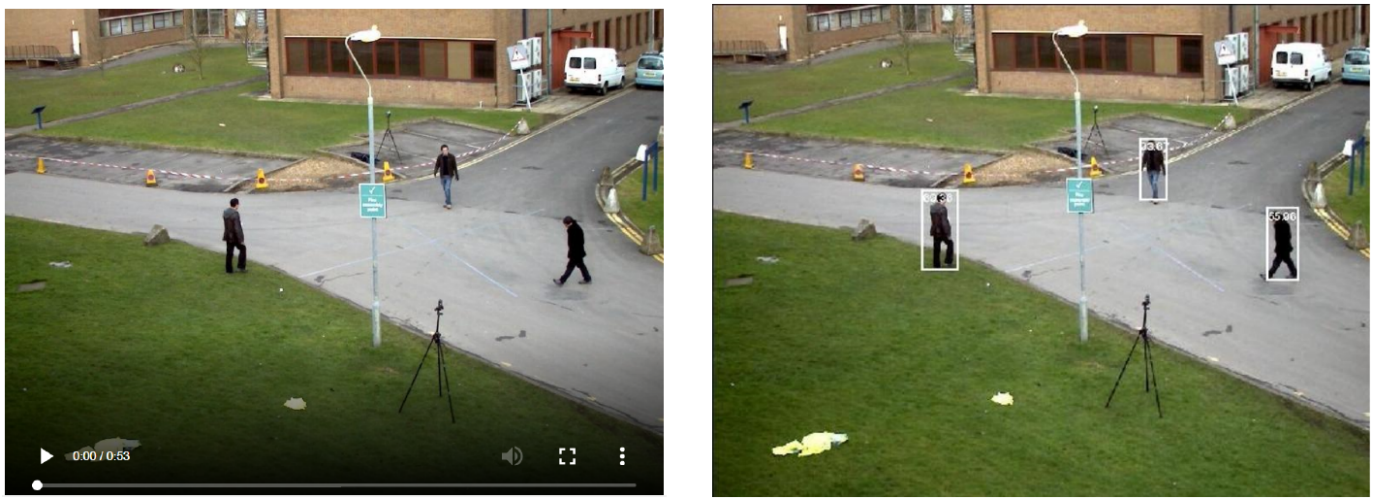

Figura 2. One frame sample from the sequence **S2.L1** in the *view-0* (left) and the same frame with the ground truth in bounding boxes (right).

Thus, it is important to have, say a gold standard, or the ground truth (GT) positions of the bounding boxes to perform a comparison between the GT bounding boxes and the ones estimated by the algorithm. To obtain the bounding boxes GT, please visit the link [Ground Truth]. This link provides a XML file containing the GT positions of the bounding boxes for each frame in the sequence. In our particular case you should download the XML file (the **complete** version) corresponding to **S2.L1 (12-34)** (in Section **PETS 2009**). For a given frame, the XML file contains the location and sizes of bounding boxes for the detected objects. The information is as follows:

- "object ID" that corresponds to a given detected object (*i.e.*, usually a pedestrian),

- "h", hight of the bounding box,
- "w", width of the bounding box, and
- "xc" and "yc", the centroid of the bounding box.

As a first stage of this work, the students are welcome to plot the GT (readable from the XML) and draw these bounding boxes in each frame, as shown in Fig. 2 right.

In the following, I will suggest some lines of work as follows:

1) Perform the tracking of pedestrians. A bounding box should be visible for each detection,
2) Plot the performed trajectories. To avoid a possible excess of the information visualisation, you can plot the trajectories dynamically,
3) If possible, try to assign a label for a given pedestrian, (or all if you want to),
4) You can also provide to the user the information regarding the map of the trajectories performed in the video,
5) In the attempt to enrich the output of your algorithm, try to establish a heatmap where the color is assigned to the number of occurrences in a given position (region) of the image,
6) Think about a way to provide optical flow of the pedestrian's motion.

IV. Evaluation Metrics

Notice that, there is no perfect algorithm. This means that, no matter the approach is adopted, there is always some failures regarding the true location of the pedestrian (*i.e.* the ground truth). For instance, a merge or split in a given bounding box that can occur. Also, some misdetections may occur as well. Thus, one way to evaluate the algorithm is to use evaluation metrics. An evaluation strategy can be done as follows:

1) The first step is to built the ground truth as already mentioned.
2) After this stage, the students are in conditions to show both the ground truth and the estimated bounding boxes provided by the developed algorithm.
3) Now, evaluation must be done. To accomplish this, the following metric is suggested:
   - Provide the success plot with increase Intersection over Union (IoU) metric. The IoU metric is defined as follows:

$$IoU = \frac{R_{det} \cap R_{rt}}{R_{det} \cup R_{rt}} \tag{1}$$

   where R_{det} is the detected region estimated by the algorithm and R_{gt} is the ground truth (manual labeled) region. The value of IoU = 1, means a perfect match, the value of IoU=0 means that the target is lost. The success plot shows the percentage of frames whose bounding box overlap ratio is higher than a given threshold. The threshold values ranges from 0 to 1, and the step can be *e.g.*, 0.05.

V. READING MATERIAL

The students are welcome to read the following papers:

[1] L. Leal-Taixe, A. Milan, I. Reid, S. Roth, and K. Schindler "MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking", arXiv 2015.

[2] J. C. Nascimento and J. S. Marques. "Performance evaluation of object detection algorithms for video surveillance". IEEE Trans. on Multimedia, vol. 8, no. 4, pp. 761-774, Aug. 2006.

[3] Luis M. Fuentes and Sergio A. Velastin, "People tracking in surveillance applications", Proceedings 2nd IEEE Int. Workshop on PETS, Kauai, Hawaii, USA, Dec. 9 2001.

Final Remark: The students must be send a zip code to the following e-mail: `jan@isr.tecnico.ulisboa.pt`. **The deadline is on May 30th, at 23h59m.**