Detection and Tracking of pedestrians in public spaces

I. Introduction

Detecting pedestrians in public spaces has become one of the core topics addressed in both computer vision and image processing communities. One of the reasons is that a large number of applications can benefit from detecting the pedestrian through time. For instance, in the fields of mobile robotics to inform Human-Robot Interaction systems, or automotive providing input to Advanced Driver Assistance Systems. Another important application is to perform the tracking of pedestrians, with the goal of collecting the trajectories. The trajectories constitute a rich source of information, since it is possible to obtain the statistics of the activities being performed in a given environment. As such, we can obtain typical trajectories that occur in a given scenario, as well as to detect atypical or suspicious behaviour, that is, trajectories that deviate from the typical trajectories. However, obtaining such algorithm for trajectory collection poses several challenges. For instance, the high variability that characterizes the pedestrians; the appearance of a pedestrian on the image that is affected by the person's pose; clothing; the atmospheric conditions that influence the illumination changes; the background clutter and occlusion. All he above issues play a role in making pedestrian detection a challenging problem to be solved.

The goal of this work is to develop an algorithm capable detect the location of the pedestrians as a way to obtain the performed trajectories. Conventional handcrafted features, that has plateaued in recent years in the image processing community, will be used for this purpose.

As a final remark, please bear in mind that, **it is expected that the output of the algorithm should be able to provide enriched 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 (i) *background*, (ii) *city center*, and (iii) *regular flow*.
2) Dataset S1: include two *walking* and one *running* sequences.

3) Dataset S2: include three *walking* sequences.
4) Dataset S3: includes the subsets *multiple flows* and *event recognition*.

In the dataset S2, there exists three subsets, denoted as $\mathbf{S2.L1}$, $\mathbf{S2.L2}$ and $\mathbf{S2.L3}$. We will concentrate on the subset $\mathbf{S2.L1}$. This subset is cataloged as a *sparse crowd* and with Level 1 of difficulty (the easiest one). The students, however, are free to use more difficult sequences if they feel like. In the $\mathbf{S2.L1}$, there exists several acquisitions, each containing a different view. The views are numbered as follows: $\mathrm{View001}$, $\mathrm{View002}$, ..., $\mathrm{View008}$. We will concentrate in the $\mathbf{View001}$ sequence. Again, the students are welcome to use another view if they feel confidence to embrace more challenging views.

Fig. 1 shows some images samples belonging to the dataset $\mathbf{S2.L1}$ in the $\mathbf{View001}$ sequence.

Figura 1. Frame samples from the dataset $\mathbf{S2.L1}$ in the $\mathbf{View001}$ sequence.

III. Ground Truth data to measure the Algorithm performance

To measure the detection performance of the algorithm, we need to have some gold standard, or ground truth (GT) of the pedestrians position, that is, a reference to conduct the comparison. One way to measure the accuracy performance is to compare the algorithm's predicted output with the GT locations. This is achieved by using bounding boxes that surrounds each pedestrian. For all the sequences mentioned in Sec. II the ground truth (GT) is provided. The GT concerns the real position of the object (*i.e.*, the pedestrians locations in the image domain) using a bounding box. Thus, one way to conduct the performance analysis is to compare the bounding boxes of the GT against the ones obtained with the algorithm. The GT information is available in [Ground Truth]. This link provides a XML file containing the following information:

- "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.

In our particular case, you should download the XML file (the complete version) corresponding to $\mathbf{S2.L1}$ (12-34) (in Section PETS 2009).

Fig. 2 shows an example of an image sample (left) and the same image with the corresponding GT detections represented in bounding boxes (right).

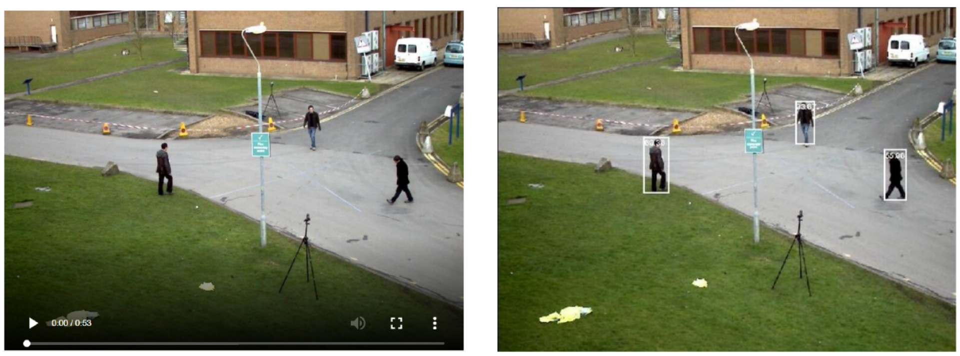

Figura 2. One frame sample from the sequence $\mathbf{S2.L1}$ in the $View001$ (left) and the same frame with the ground truth in bounding boxes (right).

The work will have several goals, the majority intend to enrich the *visual information* that can be extracted from the image sequence. Thus, the students are welcome to fulfill the following challenges:

1) Plot the GT (readable from the XML) and draw these bounding boxes in each frame, as shown in Fig. 2 right. (3.0v)
2) Now, using your detector algorithm, perform the tracking of pedestrians. The predicted bounding boxes should be visible for each detection. (4.0v)
3) Plot the performed trajectories. To avoid a possible excess of the information visualisation, you can plot the trajectories dynamically. (4.0v)
4) Assign a different label for each pedestrian. The labels should be consistent through time. (2.0v)
5) Provide to the user, the information regarding the map (*i.e.* occupancy) of the trajectories performed in the video. Specifically,
   - Provide a heatmap, using a Gaussian distance metric, where the color is assigned to the number of occurrences in a given position (region) of the image. Concerning this regard different heatmaps can be generated, this can include (i) static heatmap, (ii) dynamic heatmap. (2.0v)
   - Generalize for other distance metrics. (1.0v)

6) Think about a way to provide optical flow of the pedestrian's motion. Implemented solutions are welcome. (1.0v)
7) Provide an evaluation performance of the algorithm. Specifically provide: (i) the success plot (see Sec. IV for details), and (ii) the percentage of False Negatives (FN) or misdetections and False Positives (FP). Please provide figures illustrating the success plot, FPs and FNs, and also some frames illustrating the FPs and FNs. (3.0v)

IV. Evaluation metrics

One important issue to be considered is that every algorithm hat its own limitations. This means that, no matter the approach is adopted, there is always some failures regarding the true location of the pedestrian. 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 above.
2) After this stage, the students are in conditions to show both the ground truth and the estimated bounding boxes provided by the algorithm.
3) Now, evaluation must be done. To accomplish this, the following metric is suggested:
   - Provide the *success* plot using the Intersection over union measure (IoU) that is defined as follows:

     $$IoU = \frac{R_d \cap R_{gt}}{R_d \cup R_{gt}} \tag{1}$$

     where R_d is the detected region estimated by the algorithm and R_{gt} is the ground truth (manual labeled) region. Basically, the IoU provides a measure of the overlap (or match) between the R_d and R_{gt}. A score of $\mathrm{IoU} = 1$ means a perfect match is obtained, and $\mathrm{IoU} = 0$, means that the target is loss.

     The success plot shows the percentage of frames whose bounding box overlap ratio is higher than a given threshold. For threshold, it can be considered the values ranging from 0 to 1, with step of, *e.g.* 0.05.

Deadline: The students must upload the projects to Fenix until, April 22th, 23h59m.

V. Reading material

The students are welcome to read the following paper:

[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.