LONG-TERM GEO-POSITIONED RE-IDENTIFICATION DATASET OF URBAN ELEMENTS SUPPLEMENTARY MATERIAL

Paula Moral, Álvaro García-Martín, José M. Martínez

Video Processing and Understanding Lab, Universidad Autónoma de Madrid, Madrid, Spain.

Fig. 1. Recording Setup

1. SUPPLEMENTARY MATERIAL

This document describes the proposed dataset, how the data was collected and organized, what kind of information contains, the ethically and responsibly possible implications, and how it will be made available and maintained. The dataset is available in a drive link on the following github https://github.com/vpulab/UrbAM-ReID

1.1. Data Collection

Data collection has been performed with a GOPRO HERO11 Black with 1920×1080 pixels spatial resolution and 24 frames per second. The camera has been placed on the hood of a vehicle as can be seen in Figure 1, After data collection, annotation has been conducted using the CVAT tool [1]. Each annotation includes attributes such as its class (currently, containers, rubbish bins or crosswalks), its identity within its class (id ReID), its position in the frame (foreground or background), and, in the case of containers, the subclasses (paper, waste, plastic, glass, clothes, or organic). Manual annotation is applied to the urban objects in the first video. To automate annotations for subsequent videos, the YOLOv7 algorithm [2] is fine-tuned with the previously annotated objects' bounding boxes (bboxes). This facilitates the generation of object detections in the remaining videos, requiring manual intervention solely for the annotation of identities, classes and foreground/background attributes, and the rectification of detection errors.

This dataset contains currently three different urban objects: waste containers, rubbish bins and crosswalks. In total it contains 4,203 bboxes of 972 identities, of which 1,912 bboxes and 320 identities are for containers, 1,778 bboxes and 495 identities for crosswalks and 157 identities and 513 bboxes for rubbish bins. Each annotated identity needs to be present in at least two recordings. Crosswalks and rubbish bins are static objects, which will only vary in position or appearance from one recording to another due to possible deterioration, vandalism, or possible improvements by the responsible entity. However, in the case of containers, as they are not fixed elements, each time the waste collection company picks up the trash, the appearance of the bins due to their position could change. Due to this situation, containers with the same subclass in the same location are annotated with the same id ReID.

The span of the dataset is five months, being the first sequence recorded in March (late winter), the second one in May (middle of spring) and the last two ones in the same day of July (summer). The last sequence of July is recorded following the same trajectory but in the opposite direction.

1.2. Data Organization

Data organization is also explained in the Readme file in the UrbAM-ReID folder. UrbAM-ReID dataset is organized as follow:

./Code

- This proposal utilizes two open-source state-of-the-art works:
  - Part-Aware-Transformer
  - ReID Strong Baseline
- This folder includes the modified codes to incorporate this dataset into each system.
- To evaluate the system, it infers the data using `update.py` (or `update_GPS.py`) and evaluates the obtained

This work is part of the tasks related to the SEGA-CV (TED2021-131643A-I00) project funded by the Ministerio de Ciencia e Innovación of the Spanish Government

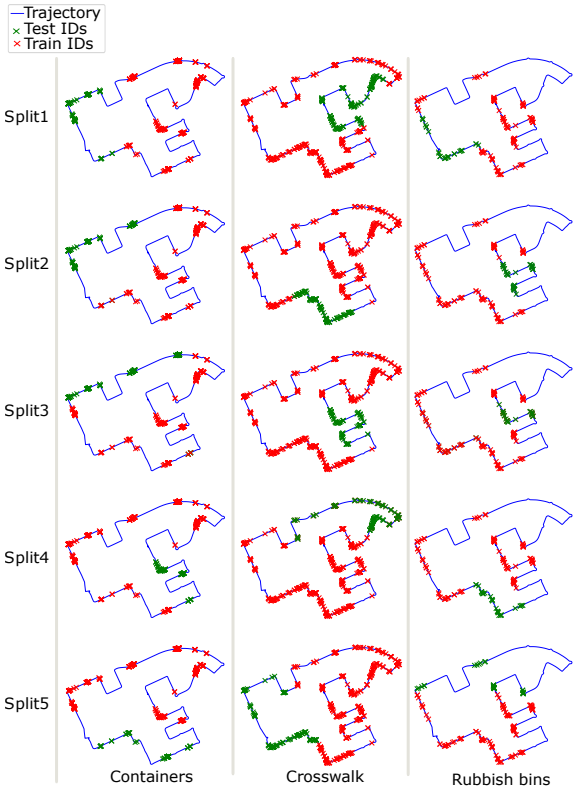

Fig. 2. Five train and test splits for each urban object.

`track.txt` results with `Evaluate_UrbAM-ReID.py`. The arguments to this update files are the same than the ones appear in githubs for the test scripts.

./GPSannotations

- Divided into four subfolders, each related to a sequence (March, May, July, and July inverse). Each subfolder contains the CSV file with GPS annotations.
- The first row is a header indicating the columns (`frame`, `lat`, `long`, `azi`).
- The subsequent rows of the file provide information on GPS coordinates per frame.
- *GPS annotation is included in this proposal as a post-processing step, but the possibilities of use are extensive.*

./ReIDSequences

- These folders contain annotated files obtained from CVAT annotations.xml with all the attributes, and the `images_output` with the format `XXXXXX_Y_Z.jpg`, where `XXXXX` is the original frame name, `Z` is the annotated ID, and `Y` is the number of this identity in this location (in case of containers that is consider the same identity per same container class in the same location, it enumerates the number of containers of the same class in this location).

./splits

- It collects five different splits per urban object (see Figure 2.
- For each split, it has baseline data (`./image_train`, `./image_test`, `image_query`).
- Then, in `./FG`, there are foreground annotated objects for the baseline.
- In `./inv`, the baseline data adds the inverse as a query (the baseline query is now in `image_test`).
- In `./inv/FG`, there are foreground annotated objects for the data including the inverse.
- Each data scenario includes the following files with annotations:
  - `train_label.xml`, `test_label.xml`, `query_label.xml`: includes cameraID (each ID=1 is May sequence, ID=2 is March, ID=3 is July, and ID=4 is July inverse). Image name related to this split, and object ID.
  - `train_label_all`, `test_label_all`, `query_label_all.xml`: includes the same cameraID; in this case, the imageName is the frame of the original sequence (the same as "frame" in GPS .csv, and in `XXXXXX` in image output, and the same as in `ReIDSequences/../annotations.xml` attribute image name). The rename is the image name in this folder (in the case of the `FG` folder, it also appears as `rename2` since it is related to the baseline folder). In `update_GPS.py` scripts there are examples of how to associate the GPS information in CSV to each files.

Figures 3 and 4 show more visual results examples, comparing the baseline set with the introduction of just the inverse, and the inverse with GPS.

1.3. Ethically, responsibly used and license

As the images used are crops of each urban object, there are no appearances of personal identities, so there would be no ethical concerns or regional legal requirements. This work is licensed under CC BY-ND 4.0 This license enables reusers to copy and distribute the material in any medium or format in unadapted form only, and only so long as attribution is given to the creator. The license allows for commercial use. CC BY-ND includes the following elements:

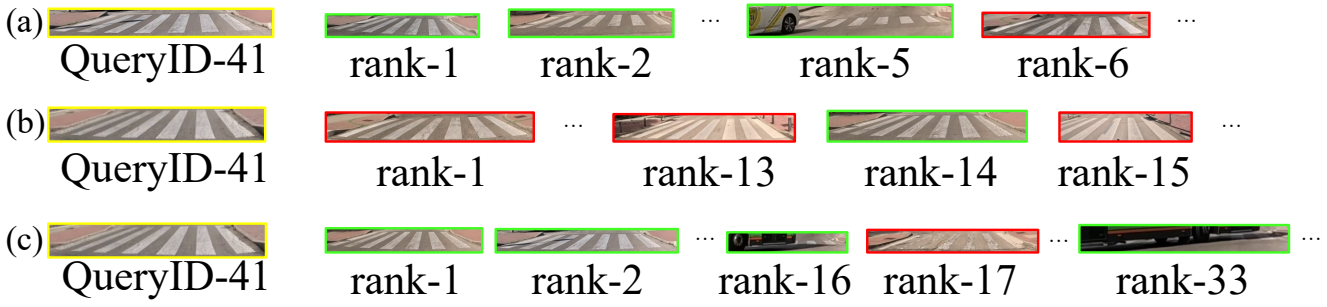

Fig. 3. Ranked results for crosswalk object in case (a) baseline set, (b) including inverse set and (c) including inverse and GPS.

Fig. 4. Ranked results for container object in case (a) baseline set, (b) including inverse set and (c) including inverse and GPS.

BY: credit must be given to the creator.

ND: No derivatives or adaptations of the work are permitted.

1.4. Mantenance Plan

For the submission the dataset is located in a Github, but for the final version it will be available in Zenodo. It is an open dissemination research data repository for the preservation and making available of research educational and informational content. It provides a unique persistent identifier (DOI) and long-term sustainability.

2. REFERENCES

[1] CVAT.ai Corporation, "Computer Vision Annotation Tool (CVAT)," Nov. 2023.

[2] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao, "Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2023, pp. 7464–7475.