

The HDA+ data set for research on fully automated re-identification systems

Dario Figueira, Matteo Taiana, Athira Nambiar, Jacinto Nascimento and Alexandre Bernardino

Institute for Systems and Robotics - Lisbon

Abstract. There are no available datasets to evaluate integrated Pedestrian Detectors and Re-Identification systems, and the standard evaluation metric for Re-Identification (Cumulative Matching Characteristic curves) does not properly assess the errors that arise from integrating Pedestrian Detectors with Re-Identification (False Positives and Missed Detections). Real world Re-Identification systems require Pedestrian Detectors to be able to function automatically and the integration of Pedestrian Detector algorithms with Re-Identification produces errors that must be dealt with. We provide not only a dataset that allows for the evaluation of integrated Pedestrian Detector and Re-Identification systems but also sample Pedestrian Detection data and meaningful evaluation metrics and software, such as to make it "one-click easy" to test your own Re-Identification algorithm in an Integrated PD+REID system without having to implement a Pedestrian Detector algorithm yourself. We also provide body-part detection data on top of the manually labeled data and the Pedestrian Detection data, such as to make it trivial to extract your features from relevant local regions (actual body-parts). Finally we provide camera synchronization data to allow for the testing of inter-camera tracking algorithms. We expect this dataset and software to be widely used and boost research in integrated Pedestrian Detector and Re-Identification systems, bringing them closer to reality.

1 Introduction

The goal of a person re-identification system is to recognize and keep track of individuals as they appear and travel in heterogeneous scenarios covered by a camera network with non-overlapping (or low overlapping) fields-of-view. Re-identification has been subject of much research in computer vision due to its usefulness in a large number of application, e.g. surveillance, smart spaces, border control, crime prevention, and robotics, to quote a few (e.g., [1,2,3]). Still, it is difficult to reliably re-identify an individual in a camera network. More specifically, when a person disappears from a given view, it is difficult to differentiate it from other targets in a different view. Difficulties arise from a multitude of factors: visual similarity between different people, occlusions, poor quality of video data and varying imaging conditions (illumination, viewing angle, distance ranges, etc.).

Most existing re-identification systems still rely, to a large extent, on human supervision and intervention. Usually a human operator is assigned to constantly monitor a large number of cameras that must be watched, interpreted and acted upon. This has, of course, several shortcomings: it is costly, inaccurate and subject to human errors. This problem motivated research on automatic systems able to effectively reduce the overhead of a human operator in attaining the goals of a re-identification system.

In the classic strategy to re-identification, examples are manually selected from the images by cropping the regions occupied by persons. First, manual crops of persons with known identity are stored in a gallery. Such gallery can be seen as a training set for the recognition of persons in other views of the camera network. Then, in novel views of the camera network, *a priori* unknown persons (probes) are selected for re-identification. Re-identification is accomplished by using some sort of matching (e.g. nearest neighbour [4], in the simplest case) between the gallery and probes. This is usually performed by learning discriminative features followed by template matching using appropriate distance measures to assess the similarity between the gallery and probe images. Most existing works address the re-identification problem giving focus on the above aspects, *i.e.* extracting local features, learning classifiers and distance measures.

Concerning local features, existing approaches typically exploit texture [5], spatial structure [4], color [6] or a combination of them [7].

Regarding distance measures, a variety of works have been published proposing different approaches, e.g. weighted L2-norm distance metric [8] or Bhattacharyya distance [4]. The distance can also be learned to provide the optimal similarity measure between a pair of person images. Several distance learning approaches are available in the literature. In [9] it is proposed a brightness transfer function. A new relative distance comparison is proposed in [10] for maximizing the probability of a pair of true matches. Principal component analysis combined with local Fisher discriminative analysis is proposed in [11] to match visual features. Supervised [12] or unsupervised [13] approaches are also proposed to extract relevant features and combining them into a single similarity function.

Other works reformulate the problem of re-identification as a ranking problem, where the potential true match is assigned with the highest rank [14] rather than a distance metric converting, in this way, the re-identification problem into a relative ranking problem.

All the above methods assume pre-cropped images of pedestrians to perform re-identification. In this paper we present a dataset and propose methodologies to relax this assumption. We want to perform re-identification in full images without requiring manual intervention in the selection of the probe data. Our aim is to move forward on the automatization of re-identification systems to facilitate the browsing of large databases of images and simplify the task of a surveillance system operator.

More concisely, we present **in Section 2** a dataset (HDA+) acquired in a heterogeneous camera network with low overlapping fields of view particularly suited to evaluate the integration of pedestrian detectors with re-identification

algorithms. The network is composed of 13 synchronized cameras of different resolutions, fields of view and perspectives. We provide **in Section 3** an evaluation framework that allows for the integration of pedestrian detectors (PD) and re-identification algorithms (RE-ID). The pedestrian detector automatically selects, from the incoming stream of images, the probe or test samples that are used as input to the re-identification algorithm, in a fully automated system (PD+REID). We provide sample data and software that allows for the evaluation of PD, RE-ID and tracking algorithms. Moreover, we provide additional features in the data set from which we highlight: (i) comparisons between classical and integrated systems (*i.e.* relying on ground truth or PD detections); (ii) evaluation metrics that take into account errors in the pedestrian detection stage; (iii) default gallery sets and pedestrian probes for easy operation; (iv) the ability to replace the default functionalities with user custom code; and (v) pre-computed body-part bounding boxes to facilitate the development of re-identification algorithms. With this dataset and tools we expect to attract newcomers and motivate established researchers to the problem of fully automated re-identification systems. **In Section 4** we focus on the integration of PD and RE-ID for which we provide sample code to evaluate the influence of different integration modalities. **In Section 5** we perform the quantitative evaluation of the proposed methodologies illustrating their advantages with respect to a naive integration. Finally **in Section 6** we conclude and suggest future developments.

2 The HDA+ data set

The dataset was acquired from 13 indoor cameras distributed over three floors of our research department, recording simultaneously for nearly 30 minutes. The video recordings exhibit a high degree of variability: image resolutions include VGA, 1MPixel and 4MPixel with corresponding frame rates of 5FPS, 2FPS and 1FPS. Moreover, the cameras were set to acquire video from different points of view, affecting the geometry of the imaged scene and resulting in different distance ranges at which people are visible in each video. The different illumination conditions of the recording areas make the dataset challenging for tasks such as person detection and re-identification. Table 2 provides an overview of the details of the video sequences.

The dataset is fully labelled. Each appearance of a person in the sequences from 13 cameras was manually annotated. The annotation information consists in the position and size of the enclosing Bounding Box (BB), the unique ID for the person, the frame number, the camera index and the occlusion flag. The occlusion flag indicates whether the person is partially occluded or fully visible. The BB's were set to fully and tightly enclose each person. For partially visible people, the extent of the full body was estimated and the Bounding Box was set to enclose it. The only exception to this rule applies to truncated people, *i.e.*, people whose projection lies partially outside the image boundaries. In that case, the bounding box was also truncated. Some of the fully visible samples



Fig. 1. Snapshots of the sequences acquired in the data set.

are provided as default gallery samples (default training set): three to five good quality images of each person, one image per appearance in each camera (a total of 250 cropped images). All other cropped images may be used as test data. This data constitutes the Ground Truth (GT) information for the dataset. A detailed documentation on the setting and data acquisition of the camera network, GT labelling procedure, and benchmarking of video analytics algorithms using the dataset can be found in the earlier publication by the authors in [15].

CAM	02	17	18	19	40	50	53
640x480	✓	✓	✓	✓	✓		
1280x800						✓	✓
2560x1600							
Frame rate	5	5	5	5	5	2	2
No of frames	9819	9897	9883	9878	9861	2227	3521
Start time	11:57:52	11:57:57	11:58:02	11:58:06	11:58:20	12:12:29	12:01:07
End time	12:30:56	12:30:56	12:30:59	12:31:02	12:31:11	12:31:02	12:30:27

CAM	54	56	57	58	59	60
640x480						
1280x800	✓	✓	✓	✓	✓	
2560x1600						✓
Frame rate	2	2	2	2	2	1
No of frames	3424	3798	3780	3721	3670	1728
Start time	12:01:31	11:59:31	11:59:36	12:00:40	12:01:06	12:02:06
End time	12:30:02	12:31:09	12:31:09	12:31:42	12:31:40	12:30:53

Table 1. camera information

In this updated version of the HDA dataset (HDAv1.1), we are releasing additional functionalities and auxiliary data. The dataset contains a general evaluation software which measures the performance of a RE-ID algorithm both in terms of the standard Cumulative Matching Characteristic (CMC) curve and in terms of a point in Precision/Recall (P/R) space. We argue that the P/R plot carries complementary information w.r.t. the CMC curve and that they should be used together to characterize the performance of a RE-ID system (see Section 3 for details).

We provide an evaluation mode that is the de facto standard for RE-ID algorithms: “MANUAL_{clean}”. In this mode the test examples are hand-cropped images of fully visible people. We also provide a more challenging evaluation

mode: “MANUAL_{all}”, in which the test set comprises all of the hand-labelled persons, including the partially visible ones. These test sets are obtained from the GT labels on the test sequences (see Section 4 for details). For the user’s convenience, we also provide the body-parts detections for all the cropped images in the test set. The parts detections are computed with the code from [16].

One of the main novelties in the current version of the dataset is the release of an architecture for a fully automated RE-ID system. We provide an integrated PD+REID system in which the input images for the RE-ID module are the detections computed by the PD module. The user of the dataset is allowed to plug in a custom PD module or take advantage of the auxiliary data we provide, in the form of the detections computed with the ACF [17] detector and the corresponding body parts detections. Integrating the PD and the RE-ID modules is not trivial because of the noise present in the detections. We provide two modules that help handle such noise. The user can choose whether to activate each module for one specific experiment, leading to four evaluation modes of the integrated system (see Sec. 4 for more details).

The synchronization of the various video sequences of a set is needed in many contexts. It is for instance essential for multi-camera tracking as it provides the user with a coordinated information of the scene. Multi-camera tracking, in turn, provides useful information for a plethora of video analytics algorithms, including person Re-Identification algorithms. We provide a complete synchronization of the video sequences with the current release of the dataset. Figure 2 and Table 2 illustrate the start time, end time and the duration of capture for each of the 13 cameras used for data acquisition. Based on the frame rates of the cameras (1, 2 and 5FPS), we selected an optimal time unit of 0.1 seconds for the synchronization framework.

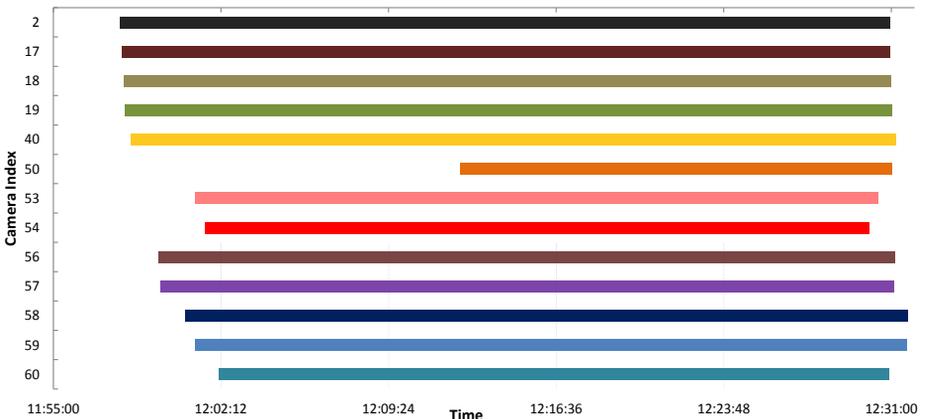


Fig. 2. The video sequences for the HDA dataset were captured for half an hour period using 13 cameras distributed in 3 floors of a building. The data acquisition was not perfectly simultaneous, being the earliest sequence frame at 11:57:52 in cam02 and the latest sequence frame at 12:31:42 in cam58.

3 Evaluation of RE-ID Systems

Most works on RE-ID consider as input pre-cropped images tightly enclosing the pedestrians to be re-identified. All training and test examples are manually selected and cropped to depict one fully visible person. The RE-ID algorithm matches test samples to the training samples and outputs a ranked list of person ID's: the lower the rank, the more confidence there is in that ID being a correct match to the detected person. Evaluation is typically done using Cumulative Matching Characteristic (CMC) curves. The CMC curve shows how often, on average, the correct person ID is included in the best K matches for each test image. The overall performance is measured by the normalized Area Under the CMC curve (nAUC).

In this work we aim at systems that are able to take full images as input, and not only pre-cropped images. One solution to tackle this problem is to use a pedestrian detector (PD) to localize persons in the video frames and send these results to a Re-ID algorithm. Such integration of a pedestrian detector with a re-identification algorithm is denoted PD+REID. From now on we concentrate on the problem of evaluating the results of PD+REID systems.

The integration of a Pedestrian Detector with a Re-Identification algorithm is not trivial, since the output of PD is prone to errors. Several types of errors may occur in the modules of the PD+REID system:

- A detected object does not correspond to a person in the image. This is termed a False Positive (FP) detection, is an error originated in the PD and it will generate a classification error at the RE-ID level because there is no correct ID for a FP.
- The system fails to detect a person. This is denoted a Missed Detection (MD) and happens whenever a ground truth Bounding Box has no corresponding detection with enough overlap. Again, this is an error of the PD module which causes a person to go unnoticed and unclassified at the RE-ID level.
- A detected object corresponds to a valid pedestrian (True Positive detection) but its rank-1 estimated class is different from the ground truth class. This is an error of the RE-ID module and we denote it an Incorrect Identification.

When a person is detected by the PD module, and is well classified by the RE-ID module we have a Correct Identification.

The CMC curve overpenalizes False Positives while ignoring Missed Detections. In order to fairly evaluate False Positives, and appreciate the effect of Missed Detections we also compute precision and recall statistics:

$$\begin{aligned}
 - \text{Precision} &= \frac{\text{Correct Identifications}}{\text{True Positive Detections} + \text{False Positive Detections}} = \frac{\text{Correct Identifications}}{\text{Number of Detections}} \\
 - \text{Recall} &= \frac{\text{Correct Identifications}}{\text{True Positive Detections} + \text{Missed Detections}} = \frac{\text{Correct Identifications}}{\text{Number of Person Appearances}}
 \end{aligned}$$

We also define two labels to take into account the possibility of re-identification algorithms able to identify persons not included in the gallery, i.e. new persons (ID-NEW), and detections not corresponding to actual people, i.e. false positives (ID-FP). In scenarios with access control it is possible to know at each time which

persons exist in the surveilled space. We denote these cases as Closed Spaces. On the other hand, there are scenarios where new persons may appear in the images (Open Spaces). If the user selects the Closed Space scenario, all persons not in the gallery set are ignored for evaluation purposes. In the Open Space scenario, persons not in the gallery set are given the special tag ID-NEW and considered in the evaluation. If the re-identification algorithm correctly identifies a person as ID-NEW, the corresponding detection counts as a Correct Identification. If a new person is attributed one of the existing ID's, then it is accounted as an Incorrect Identification. Analogously, if detections not matching the ground truth enough are passed to the RE-ID algorithm and it is able to understand that the sample is not a person, it assigns the label ID-FP and the evaluation function considers it a Correct Identification. If the label ID-FP is associated to a valid person, it is taken as an Incorrect Identification.

3.1 Using the Evaluation Tool

Here we shall describe with some detail the software tool provided. The general view of the evaluation system is shown in Fig. 3. The input to the evaluation system is a list containing information about all detected objects in the test sequences and corresponding ranked list of person ID's. Each item of the input list corresponds to a detection and contains the following data:

- Camera number
- Frame number
- Localization and extent of a detection (Bounding Box information)
- List of ranked ID's

The evaluation function checks this data against the ground truth and computes the CMC curves and Precision-Recall statistics. The same data format is used to perform the classical RE-ID algorithm evaluation (RE-ID performed on manually cropped images), by using the ground truth Bounding Boxes instead of automatic detections.

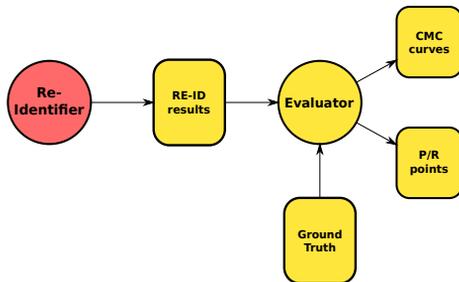


Fig. 3. The evaluation tool receives the computed detections and RE-ID ranks and produces the CMC and Precision/Recall values by comparison with the ground-truth information.

While defining an evaluation experiment the user needs to specify:

- two disjoint sets of cameras, one for training and the other for testing;
- the source of the training data: the default training set provided with the data set or a user-provided one;
- the source of the test data: the ground truth hand-labelled Bounding Boxes (BB), the BB’s obtained by the the automatic pedestrian detector, or BB’s manually provided by the user;
- some options on whether and how to filter the test data BB’s (as described in Section 4.2).

When using the test data provided in HDA+, the user may employ one of three different test sets: (i) non-occluded ground truth test samples, (ii) the whole set of ground truth test samples (including the occluded ones), (iii) or a set of test samples obtained by the ACF pedestrian detector [17]. We term the evaluation modes related to the use of these test data subsets, respectively as: (i) `MANUALclean`; (ii) `MANUALall`; and (iii) `DIRECT`. In **Section 5** we show an experimental evaluation of these modes.

Furthermore the user should select whether to consider the special tags `ID-NEW` and `ID-FP` to correctly compute the evaluation scores according to the algorithms capabilities.

At the current state of the evaluation system, we do not allow the possibility of dynamically growing the gallery as new persons are identified in the test set. This is an important point to consider in future work.

4 Evaluation of Integrated PD+REID System

Most works on the Re-Identification (RE-ID) of people operate on manually cropped images. However, in order to achieve a fully automated RE-ID system, the automatization of the detection is needed, i.e., the cropping must be performed by a Pedestrian Detection (PD) algorithm.

The integration of PD and RE-ID poses several challenges. The False Positives (FP’s) and Missed Detections (MD’s) generated by the PD have an impact on the performance of the whole system: FP’s lead to cropped images that are impossible to correctly associate to the ID of a person, while MD’s do not generate a cropped image at all, making the identification of a missed person impossible.

Moreover, even true positive detections rather than manually selected Bounding Boxes can give rise to a more difficult RE-ID problem. First, the alignment and size of a detection bounding box with respect to the detected person are bound to be less precise than the regions selected by a user. Second, common test sets for person RE-ID consist exclusively of fully visible people, while detections can match people imaged under varying degrees of occlusion.

Connecting a PD system directly to a RE-ID one yields poor results because of the aforementioned sources of noise. In previous work [18] we introduced two improvements to the naive integration scheme: the Occlusion Filter and the use of a False Positive Class. The Occlusion Filter is a processing block which lies

between the PD and RE-ID systems (see Fig. 5), its goal is to reduce the incidence of partially occluded detections in the data fed to the RE-ID system. The use of a False Positive Class stems from the observation that False Positive detections (the ones which do not correspond to a person in the image) are impossible to be correctly classified by the RE-ID system. Defining a FP class means explicitly modelling the typical FP's of a data set. This makes a correct classification of FP detections possible, which in turn allows for a coherent evaluation of the performance of the integrated system. In the current release of the HDA dataset, we distribute code and auxiliary data so that the researchers in the RE-ID field can evaluate their algorithms in different working conditions: we provide both manually cropped images and crops obtained by applying a state-of-the-art PD [17]. We provide the code that implements the Occlusion Filter and that builds the False Positive Class. All data and modules are part of an intuitive software architecture which makes it easy to integrate the user's RE-ID system and evaluate its performance in different experiments.

4.1 Naive integration of PD and RE-ID modules

In the classical evaluation of RE-ID systems, the input consists of a set of manually cropped images, associated with their respective person ID's. In the integrated PD+REID system a Pedestrian Detector is used to generate the cropped images which are presented as input to the RE-ID system.

In the current release of the dataset, we provide detections obtained with a state-of-the-art PD system, the ACF detector [17]. Users of the dataset can test their RE-ID algorithms in an integrated system without the need of implementing a detector. Another possibility is to plug another detector in the provided architecture, and simply select which one to use in each experiment, via the main evaluation script (please see the documentation for more details¹).

4.2 Occlusion Filter

We showed in [18] that filtering the detections produced by a PD system based on geometrical reasoning positively affects the performance of the integrated PD+REID system. We provide the implementation of the Occlusion Filter (OF) with the HDA+ data set. The Occlusion Filter is a block which aims at filtering the detections prior to sending them to the RE-ID system, rejecting the ones which appear under a significant degree of occlusion (e.g., the woman in Fig. 4). Such detections are particularly hard to classify because the features computed on them can be a mixture of those generated by the occluding and the occluded person. The visibility information of overlapping detections is not computed by the detection system, it is inferred instead with a heuristic based on scene geometry: in a typical surveillance scenario, with cameras mounted above the head of people and pointing downwards, the closer a person is to the camera, the lower its projection will reach on the image (see the arrows in Fig. 4).

¹ <http://vislab.isr.ist.utl.pt/hda-dataset/>



Fig. 4. Example of two overlapping detections and determination of which is occluded. The two Bounding Boxes have a high degree of overlap as shown by the yellow area in the image. In a pair of overlapping detections the one with a higher positioning of its Bounding Box lower boundary (see the arrows) is the occluded one based on geometric considerations. Thus in this image, the woman is heavily occluded.

We provide the code for the Occlusion Filter with the data set. The user can choose whether to enable it for each experiment (see details in the documentation). When the OF is active, the user needs to select the minimum amount of overlap at which we reject the occluded detection. Furthermore, the architecture allows for the user to substitute the Occlusion Filter with a different kind of filter, without compromising the functionality of the rest of the evaluation code.

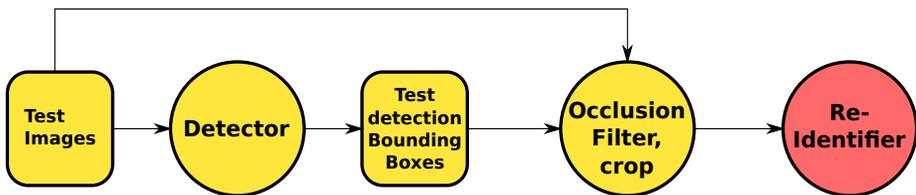


Fig. 5. Block diagram of the Detection subsystem employed in the evaluation of an integrated PD+REID system. First, a detector is run on the test images. Then, the resulting Bounding Boxes are filtered by the Occlusion Filter and the corresponding crops are generated. Such crops form the input data for the RE-ID system.

4.3 False Positives Class

We also showed in [18], that integrating a PD and a RE-ID module without appropriately managing the FP's produced by the PD, leads to inconsistencies in the evaluation. Each FP example produces an incorrect classification, leading

to CMC curves which do not reach 100% accuracy (see the green line in Fig. 8). Introducing a False Positive class allows for the correct classification of a part of the FP detections and for well behaved CMC curves.

We include the code that builds the FP class in the current release of the HDA+ data set (see Fig. 6). FP examples are harvested running the detector on all images of the specified training set. Each detection with zero overlap with any Ground Truth Bounding Box is classified as a FP. In a real scenario, this would be done by computing the detections when no persons are present in the environment and creating the FP class with these detections. The set of FP's is used by the RE-ID system to create a model for the FP class. Note that the number of harvested FP's tends to be higher than the usual number of training samples for a given pedestrian.

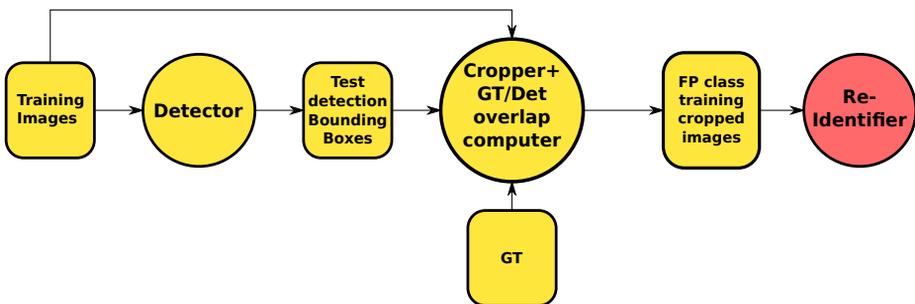


Fig. 6. Block diagram of the subsystem that computes the examples for the False Positive class. A detector is run on the training images. The detections which do not overlap at all with Ground Truth labels are deemed as False Positives. The cropped images corresponding to the False Positives are used to build a model in the RE-ID system.

5 Experimental Results

In this section we present the results obtained by applying a particular re-identification algorithm on the dataset and evaluating with the different modalities discussed in Section 3 and 4. The RE-ID algorithm is based on the Nearest Neighbour classifier, Bhattacharyya distance and HSV color histograms [6]. First we compare the RE-ID results obtained using the two manual modes, then we compare such results with those obtained by the integrated PD+REID system.

The experiments performed use the default gallery samples of cameras 50 to 59 for training, whereas camera 60 is used for testing. The annotation for camera 60 comprises 1182 Bounding Boxes, 1097 of which are of people that appear both in camera 60 and in some of the training cameras. Since we consider the Closed Space assumption in this example, the test set for the $\text{MANUAL}_{\text{all}}$ evaluation modality comprises such 1097 BB's. For the $\text{MANUAL}_{\text{clean}}$ modality, we use only

the fully visible pedestrians in the test set: 467 out of 1097. For the experiments with the PD+REID system, the ACF detector produces 2579 Bounding Boxes in the test video sequence (camera 60), of which 1167 are False Positives. In the evaluation modes in which the Occlusion Filter is active, 233 detections are filtered out, leaving a total of 2309 BB's.

The results are visualized in Figure 8 (CMC curves) and Figure 7 (P/R points). Table 2 lists the corresponding numerical values: the precision and recall statistics for rank 1, including the F-score (harmonic mean of precision and recall), while CMC curves are summarized by the values of the rank 1 point and those of the normalized area under the curves (nAUC).

Comparing the $\text{MANUAL}_{\text{all}}$ with the $\text{MANUAL}_{\text{clean}}$ experiment allows us to measure the difference in performance caused by the introduction of partially occluded exemplars in the test set. The $\text{MANUAL}_{\text{clean}}$ and $\text{MANUAL}_{\text{all}}$ baseline cases perform as expected. $\text{MANUAL}_{\text{clean}}$ receives the cleanest possible input (only fully visible pedestrians) and exhibits the highest precision of all experiments. $\text{MANUAL}_{\text{all}}$, on the other hand, receives Bounding Boxes for all the pedestrian appearances (including the ones affected by partial visibility) and reaches the highest values for recall. The CMC curve of $\text{MANUAL}_{\text{clean}}$ outperforms that of $\text{MANUAL}_{\text{all}}$ for all ranks.

Comparing the two manual modes with the naive integration of PD and REID (the DIRECT mode) we observe a drop in performance for the PD+REID system (see Fig. 8). This loss is mostly due to the fact that False Positive detections are always misclassified in the DIRECT mode. The CMC curve is thus limited in the highest accuracy value it can reach. Such loss is mitigated when we consider the two integration modules discussed in the previous section, the FP class and the Occlusion Filter. Figure 8 highlights the obvious improvement provided by using the FP class. This is due to two factors: (i) Adding a FP class allows all detections to be correctly identified at some rank, enabling the CMC curve to reach 100% accuracy; (ii) In this experiment, most of the False Positives are quite easy to re-identify, as they are generated by static objects in the scene (i.e, doors, fire extinguishers). This causes the low-rank part of the curve to be even higher than that of the manual modes. Note that we define the precision and recall statistics so that they are not affected by the quality of re-identification in the FP class.

Furthermore, we can see the improvement afforded by the Occlusion Filter when the FP class is ON (in Fig. 8 the orange dash-dot curve is always higher than the blue dotted one). As reported in Table 2, the results are as expected. By filtering out difficult to re-identify cases, the precision rises, but on the other hand, by reducing the number of detections considered, the recall lowers. When the FP class is OFF, comparing between the "DIRECT" vs "FP OFF, OCC ON" modes, we do not see any increase in accuracy or precision, possibly because the general values for this statistics are so low that the increase afforded by the OF is not significative.

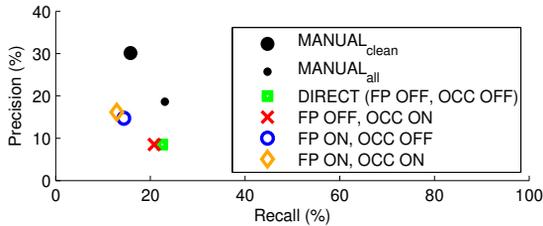


Fig. 7. Visualization of the precision-recall values as per Table 2.

Exp.	# BB's	Precision (%)	Recall (%)	Fscore (%)	1st Rank (%)	nAUC (%)
MANUAL _{clean}	462	30.1	15.8	20.7	18.7	90.6
MANUAL _{all}	1097	18.6	23.1	20.6	31.2	82.4
DIRECT (FP OFF, OCC OFF)	2542	8.5	22.5	12.4	5.2	29.1
FP OFF, OCC ON	2309	8.5	20.8	12.1	5.4	27.8
FP ON, OCC OFF	2542	14.7	14.4	14.6	62.3	91.3
FP ON, OCC ON	2309	16.2	12.9	14.4	66.7	93.5

Table 2. Statistics for the different evaluation modes presented in the text. We list the number of Bounding Boxes (# BB's) processed in each case, and report the results in terms of **Precision**, **Recall**, **Fscore**, the CMC curves' **1st rank** and its normalized area (**nAUC**). Note that we define the precision and recall statistics so that they are not affected by the quality of re-identification in the FP class, while CMC is affected by them. This leads to the very high **1st Rank** values for the modes with FP class turned ON. These values are of little practical interest. The precision-recall values can be visualized in Fig. 7.

6 Conclusions

Current re-identification algorithms are still far from practical applications due to several factors. One of them is related to the fact that typical evaluation methodologies consider the re-identification algorithm working with ideal data, in isolation from other components of an automated system. For most practical purposes, re-identification algorithms should operate in tandem with person detectors in order to seamlessly operate using conventional video data. In this paper we proposed a few contributions towards this goal. First, a few methods were developed to improve a naive integration of pedestrian detectors and re-identification algorithms. Second, a multi-camera fully labeled dataset and associated tools was developed to facilitate the evaluation of such methods. Finally some limitations on classical evaluation metrics were identified and new ones proposed. Both the dataset and the evaluation tools are available for the scientific community. We hope these can drive interest and progress in the field. In future work we aim at developing methodologies for taking into account dy-

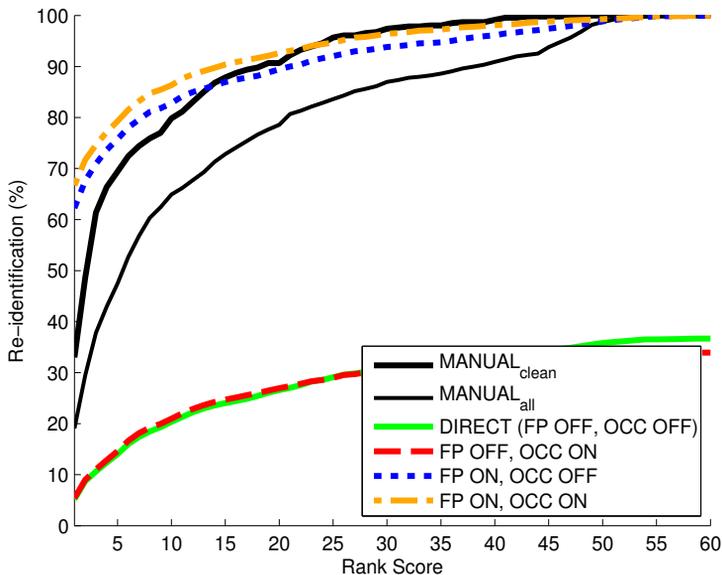


Fig. 8. CMC curves of the six Re-Identification experiments with the experimental set-ups described in the text.

namically changing galleries, according to people entering or leaving the shared space.

7 Acknowledgments

This work was partially supported by the FCT project [PEst-OE/EEI/LA0009/2013], the European Commission project POETICON++ (FP7-ICT-288382), the FCT project VISTA [PTDC/EIA-EIA/105062/2008], the project High Definition Analytics (HDA), QREN - I&D em Co-Promoção 13750, Matteo Taiana was supported by the FCT doctoral grant [SFRH/BD/43840/2008] and Dario Figueira was supported by the FCT doctoral grant [SFRH/BD/48526/2008].

References

1. Doretto, G., Sebastian, T., Tu, P., Rittscher, J.: Appearance-based person reidentification in camera networks: problem overview and current approaches. *Journal of Ambient Intelligence and Humanized Computing* **2**(2) (2011) 127–151
2. S. Gong, M. Cristani, S.Y.: *Person Re-Identification*. Springer (Jan. 2014)
3. Vezzani, R., Baltieri, D., Cucchiara, R.: People reidentification in surveillance and forensics: A survey. *ACM Comput. Surv.* **46**(2) (December 2013) 29:1–29:37
4. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: *CVPR*. (2010)

5. Madden, C.S., Cheng, E.D., Piccardi, M.: Tracking people across disjoint camera views by an illumination-tolerant appearance representation. *Mach. Vis. Appl.* **18**(3-4) (2007) 233–247
6. Figueira, D., Bernardino, A.: Re-identification of visual targets in camera networks: a comparison of techniques. *Image Analysis and Recognition* (2011) 294–303
7. Gheissari, N., Sebastian, T., Hartley, R.: Person reidentification using spatiotemporal appearance. In: *CVPR*. (2006)
8. Layne, R., Hospedales, T., Gong, S.: Person re-identification by attributes. In: *BMVC*. (2012) 24.1–24.11
9. Porikli, F.: Inter-camera color calibration using cross-correlation model function. In: *ICIP*. (2003)
10. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: *CVPR*. (2011)
11. Pedagadi, S., Orwell, J., Velastin, S., Boghossian, B.: Local fisher discriminant analysis for pedestrian re-identification. In: *CVPR*. (2013)
12. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: *ECCV*. (Jul 2008)
13. Bashir, K., Xiang, T., Gong, S.: Feature selection on gait energy image for human identification. In: *ICASSP*. (2008)
14. Prosser, B., Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by support vector ranking. In: *BMVC*. (2010) 21.1–11
15. Nambiar, A., Taiana, M., Figueira, D., Nascimento, J., Bernardino, A.: A multi-camera video data set for research on high-definition surveillance. *International Journal of Machine Intelligence and Sensory Signal Processing, Special Issue on Signal Processing for Visual Surveillance*, Inderscience Journal (InPress 2014)
16. Andriluka, M., Roth, S., Schiele, B.: Pictorial Structures Revisited: People Detection and Articulated Pose Estimation. *CVPR* (2009)
17. Dollár, P., Appel, R., Belongie, S., Perona, P.: Fast feature pyramids for object detection. *PAMI* (2014)
18. Taiana, M., Figueira, D., Nambiar, A., Nascimento, J., Bernardino, A.: Towards fully automated person re-identification. *VISAPP* (2014)