Towards Fully Automated Person Re-identification

Matteo Taiana, Dario Figueira, Athira Nambiar, Jacinto Nascimento and Alexandre Bernardino
Institute for Systems and Robotics, IST, Lisboa, Portugal

Keywords: Re-Identification, Pedestrian Detection, Camera Networks, Video Surveillance.

Abstract: In this work we propose an architecture for fully automated person re-identification in camera networks. Most works on re-identification operate with manually cropped images both for the gallery (training) and the probe (test) set. However, in a fully automated system, re-identification algorithms must work in series with person detection algorithms, whose output may contain false positives, detections of partially occluded people and detections with bounding boxes misaligned to the people. These effects, when left untreated, may significantly jeopardise the performance of the re-identification system. To tackle this problem we propose modifications to classical person detection and re-identification algorithms, which enable the full system to deal with occlusions and false positives. We show the advantages of the proposed method on a fully labelled video data set acquired by 8 high-resolution cameras in a typical office scenario at working hours.

1 INTRODUCTION

This paper tackles the problem of person re-identification (RE-ID) in camera networks. Given a set of pictures of previously observed persons, a practical RE-ID system must locate and recognise such people in the stream of images flowing from the camera network. Its classical applications are in the field of video surveillance and security systems as well as in human-machine interfaces, robotics, gaming and smart spaces. In this work we consider a set of cameras with low overlap covering our research institute facilities. When a person enters the space or passes on some key locations where her/his identity can be verified (face recognition, access control), pictures are acquired and stored in a gallery, associated to the respective identity. Such gallery provides the training data to develop methods to recognise the person on the other images of the camera network. The full RE-ID system can thus be used to update the location of a person on the covered space along time, supporting research on several topics of our interest (modelling activities, mining physical social networks, human-robot interaction).

The problem we tackle is very challenging due to a multitude of factors. The placement of a camera influences the perspective and the amount of occlusion a person is observed under, and the range of distances people are imaged at, eventually defining their height in the image. Different illumination and camera optics change the brightness and colour characteristics of the measurements. Finally, persons may change

Figure 1: A typical re-identification algorithm is based on a gallery set: a database that contains the persons to be re-identified at evaluation time. People detected in other images (probes) are matched to such database with the intent of recognising their identities. Classically, re-identification algorithms are evaluated with manually cropped probes. In this work we study the effect of using automatic probe detection in the full re-identification system.

their clothing and other appearance traits along time (possibly for disguise purposes), making the RE-ID problem particularly difficult even for humans.

Most recent state-of-the-art algorithms on RE-ID are developed and evaluated as a matching problem. The appearance information of persons stored in the gallery (training set) is matched against manually cropped images of persons in the network cameras (probe data). See Fig. 1 for an illustration of the process. However, in most RE-ID applications of interest, it is necessary to detect the probe person's bounding boxes (BB) in an automated way. This is com-

Taiana M., Figueira D., Nambiar A., Nascimento J. and Bernardino A..
Towards Fully Automated Person Re-identification.
DOI: 10.5220/0004682301400147
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 140-147
ISBN: 978-989-758-009-3

monly accomplished using one of two approaches: background subtraction and pattern recognition. We focus on methods based on the latter, because of their applicability to a wider range of scenarios, including moving/shaking cameras and single frame images. There are two main classes of pedestrian detection (PD) algorithms: monolithic or holistic (human as a whole) (Dollár et al., 2010) and part-based (human as a composition of parts) (Girshick et al., 2011). In this work we use our implementation (Taiana et al., 2013) of a monolithic algorithm based on (Dollár et al., 2010) for PD and the algorithm described in (Figueira et al., 2013) for RE-ID.

Both in background subtraction and in pattern recognition methods, the output of PD is subject to several forms of "noise", in particular missed detections, false positives and BB's misaligned with the detected people. When a RE-ID algorithm is combined with an automatic PD algorithm, the performance of the former will suffer from the imperfections of the data presented on its input. Furthermore, it is worth noticing that a class of correct detections is particularly difficult for a RE-ID algorithm to handle: the detections of partially occluded people. Currently very few works address these problems.

To leverage the combination of PD and RE-ID algorithms in a fully automated RE-ID system, we propose two important extensions. On one hand, the PD output is used to filter detections with a large degree of occlusion, which most likely contain a large amount of data not related to the detected person. Doing so prevents mis-identifications at the RE-ID stage. On the other hand, the RE-ID module is trained to directly represent a class of false positives commonly detected in the camera network. This minimises the false associations of probe "noise" to actual persons in the gallery. We evaluate our system on a recently developed data set of high-definition cameras. Eight cameras covering a large space of our research institute were used to acquire data simultaneously for thirty minutes during rush hour (lunch time). Images were fully labelled with BB locations, persons' identities and occlusion and crowd flags. We show that the proposed add-ons to PD and RE-ID algorithms are important features in a fully automated RE-ID system.

2 RELATED WORK

The problem of re-identification entails the recognition of people traversing the field of view of different non-overlapping sensors. State-of-the-art methods regarding this topic are often characterized by differences in the contextual knowledge, the learning methodology, the feature extraction, the data association and the type of gallery.

Approaches relying on *contextual knowledge* from surrounding people using human signature (Zheng et al., 2009), or attribute weighting features (Liu et al., 2012a) were proposed. Another class of approaches was developed to deal with *pose variation* (Bak et al., 2012) using Mean Riemannian Covariance patches computed inside the bounding box of the pedestrian. Shape and appearance context to model the spatial distributions of appearance relative to body parts (Wang et al., 2007), triangular graph model (Gheissari et al., 2006) and part-based models (Corvee et al., 2012; Cheng et al., 2011) constitute valuable alternatives handling pose variability. Methodologies using *supervised* learning, were also suggested. They usually include discriminative models such as SVM and boosting for feature learning (Gray and Tao, 2008; Prosser et al., 2010; Zheng et al., 2011), in the attempt to iteratively select the most reliable set of features. Another direction concerns the learning of task-specific distance functions with metric algorithms (Zheng et al., 2011; Mignon and Jurie, 2012; Hirzer et al., 2012; Li and Wang, 2013; Li et al., 2012). For all these methods, training samples with identity labels is mandatory. Approaches focusing on direct distance metrics were also proposed in the re-identification context. Symmetry-Driven Accumulation of Local Features was proposed in (Farenzena et al., 2010). In (Ma et al., 2012) the BiCov descriptor was presented, combining Gabor filters and covariance descriptors to account for both illumination and background changes. In (Liu et al., 2012a) it was shown that certain appearance features can be more important than others and selected online. In (Mogelmose et al., 2013a) it was shown that using RGB, depth and thermal features in a joined classifier are able to improve the re-identification performance. Feature extraction for *fast matching* was also investigated. In (Hamdoun et al., 2008) a KD-tree was used to store per-person signatures. Random forests were proposed in (Liu et al., 2012b) to weight the most informative features, while code-book representations were used in (Jungling et al., 2011). Person re-identification can be formulated as a *data association* problem, in matching observations from pairs of cameras. Several methods were proposed to learn this association space. Large-Margin Nearest Neighbor (with Rejection) was proposed (Dikmen et al., 2011) to learn the most suitable metric to match data from two distinct cameras, while other metrics include Probabilistic Relative Distance Comparison (Zheng et al., 2011). Rank-loss optimization was used to improve accuracy in re-identification (Wu et al., 2011) and a variant of Locally Preserving Projections (Harandi et al., 2012) was formulated over a Riemmanian manifold. Recently Pairwise Constrained Component Analysis was proposed for metric learning tailored to address the scenarios with a small set of examples (Mignon and Jurie, 2012). Also (Pedagadi et al., 2013) suggested a metric learning for re-identification, where a Local Fisher Discriminant Analysis is defined by a training set. Most of the aforementioned works use the VIPeR (Gray and Tao, 2008), ETHZ (Ess et al., 2007) and i-LIDS (Zheng et al., 2009) data sets, focusing only on the matching problem, thus neglecting the automatic detection of people.

Methods that actively integrate pedestrian detection and re-identification are still scarce in the literature. Yet, few examples are available. In (Corvee et al., 2012; Bak et al., 2012), pedestrian detection and re-identification are used, but not actually integrated. The above frameworks rely on the assumption that the people must be *accurately* detected. The work that relates the most with ours is that introduced in (Mogelmose et al., 2013b). In that work, the system full flow (i.e. pedestrian detection and re-identification) is presented with a transient gallery to tackle open scenarios. However, important issues such as how re-identification performance is penalized when pedestrian detection or tracking failures exist are not explored. The goal of our paper is precisely to explore how to enhance the link between pedestrian detection and re-identification algorithms to improve the overall performance.

3 ARCHITECTURE

For the algorithms in the state of the art, the data for the Re-Identification (RE-ID) problem is provided in the shape of hand-cropped Bounding Boxes (BB), rather than in the shape of full image frames. Such BB's are centered around fully visible, upright persons. The data is usually partitioned into training and test set and the focus of the RE-ID algorithms is on feature extraction and BB classification. However, the purpose of an automated RE-ID system is that of re-identifying people directly in images, without requiring manual intervention to produce the BB's. The natural candidate for substituting the hand-made annotation process is a Pedestrian Detection (PD) algorithm, which takes one image as input and produces a set of BB's as output. The resulting architecture is depicted in Figure 2.

Integrating PD and RE-ID poses several issues. Detecting people in images is a hard task, in fact even the best detectors in the state of the art are subject to the production of at least two types of errors: False Positives (FP) and Missed Detections (MD). Such errors have an impact on the performance of the compounded system: FP's generate BB's which is impossible for the system to correctly classify as one of the persons in the training set. MD's, on the other hand, cause an individual to simply go undetected, and thus, unclassified. Even the correctly detected persons may give rise to some difficulties: (1) the PD algorithm can generate a BB not centred around the person or at a non-optimal scale – this might hinder the feature extraction phase, prior to the classification, (2) the detected person may be partially occluded, yet again hampering feature extraction, and finally (3) there can be the case of detecting people who are not part of the RE-ID training set, posing an issue similar to that of FP's: there is no correct class that the system can attribute to them. To limit the complexity of the problem, we constrain the scenario with a closed-space assumption: we require the access to the surveilled area to be granted exclusively to people listed in the training set. In the rest of this section we propose an architecture which integrates the PD and RE-ID stages, solving some of the aforementioned issues.

3.1 Occlusion Filter

We devise the Occlusion Filter – a filtering block between the PD and the RE-ID modules – with the intent of improving the RE-ID performance. The Occlusion Filter uses geometrical reasoning to reject BB's which can harm the performance of the RE-ID stage, the harmful BB's being the ones depicting partially occluded people. A BB including a person appearing under partial occlusion generates features different from a BB including the same person under full visibility conditions. When the partial occlusion is caused by a second person standing between the camera and the original person, the extracted features can be a mixture of those generated by the two people, making the identity classification especially hard (see illustration in Figure 3). For this reason, it would be advantageous for the RE-ID module to receive only BB's depicting fully visible people. We define a first, aggressive, operation mode for the Occlusion Filter which rejects all the detections which overlap with others, the "RejectAll" mode. Though the visibility information is not available to the system, it can be estimated quite accurately with a heuristic based on scene geometry: in a typical scenario the camera's perspective projection makes pedestrians closer to it extend to relatively lower regions of the image.

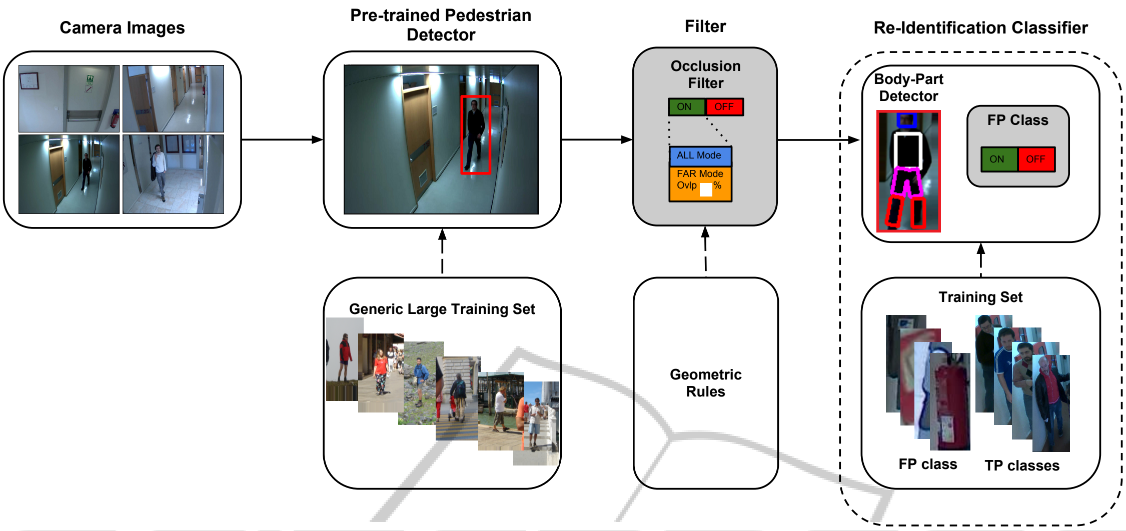
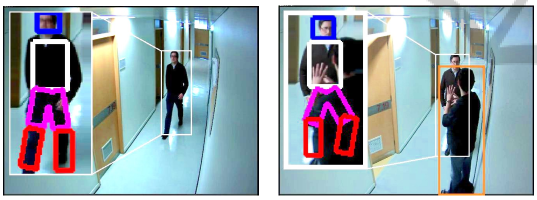


Figure 2: Architecture of the proposed fully automated Re-Identification system. The images acquired by a camera network are processed by a Pedestrian Detection algorithm to extract candidate Bounding Boxes. The Bounding Boxes are optionally processed by the Occlusion Filter, which can operate either in the “RejectAll” or the “RejectFarther” mode (see explanation in text). Lastly, the Re-Identification module computes the features corresponding to each Bounding Box and classifies it. The classification can optionally take into account a “False Positive” class.



(a) (b)

Figure 3: Example of body part detection for feature extraction in two instances: (a) a person appearing with full visibility and (b) under partial occlusion, with detection Bounding Boxes overlap, and feature extraction on the occluded person mistakenly extracting features from the occluding pedestrian. The contrast of both images was enhanced for visualization purposes.

Thus, we design the “RejectFarther” operation mode for the Occlusion Filter. In this mode the filter computes the overlap among all pairs of detections in one image and rejects the one in each overlapping pair for which the lower side of the BB is higher (as illustrated in Figure 4). Considering the mismatch between the shape of the pedestrians’ bodies and that of the BB’s, it is clear that an overlap between BB’s does not always imply an overlap between the corresponding pedestrians’ projections on the image. We define an overlap threshold for the “RejectFarther” mode of the filter, considering as overlapping only detections whose overlap is above such threshold. The impact of the overlap threshold on the RE-ID performance is analysed in the following section.

Figure 4: An example of geometrical reasoning: two detection Bounding Boxes overlap. The comparison between the lower sides of the two Bounding Boxes leads to the conclusion that the person marked with the red, dashed Bounding Box is occluded by the person in the green, continuous Bounding Box.

3.2 False Positives Class

The second contribution of this work is to adapt the RE-ID module so that it can deal with the FP’s produced by the PD. The standard RE-ID module cannot deal properly with FP’s: each FP turns into a wrongly classified instance for the RE-ID. Observing that the appearance of the FP’s in a given scenario is not completely aleatory, but is worth modelling (see Fig. 5), we introduce a FP class for the RE-ID module. In these conditions a correct output for when a FP is presented on the RE-ID’s input exists: the FP class. This change allows us to coherently evaluate the performance of the integrated system.

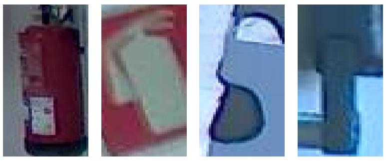


Figure 5: Example False Positive samples in the False Positive Class training set.

4 EXPERIMENTAL RESULTS

We work with the recently proposed high-definition data set described in Section 1. For our experiments we consider the closed-space assumption. A set of images with pedestrians authorised in the surveilled space is collected in a training stage and stored in a gallery set associated to the pedestrians identities. In our experiment we simulate the closed-space assumption selecting the best images of 7 out of the 8 cameras sequences for training, and using the last sequence as a test set. The training set is built hand-picking one detection per image sequence, per pedestrian, leading to a total of 230 detections for 76 pedestrians. The False Positive (FP) class is built with the detections from the training sequences that have no overlap with a Ground Truth (GT) Bounding Box (BB), for a total of 3972 detections. The test image sequence contains 1182 GT BB's, centered on 20 different people. Such people are fully visible in 416 occurrences and appear with some degree of occlusion by other BB's or truncated by the image border in 766 occasions.

Pedestrian Detection. In this work we use a state-of-the-art PD system: our implementation (Taiana et al., 2013) of Dollár's Fastest Pedestrian Detector in the West (Dollár et al., 2010) (FPDW). This module generates 1182 detections on the test camera sequence. The initial detections are filtered based on their size, removing the ones whose height is unreasonable given the geometric constraints of the scene (under 68 pixels). This rejects 159 detections and allows 1023 of them pass. Considering that three pedestrians who appear in the test set are not present in the training set, we remove the corresponding 59 detections in the test set from the detections' pool. This leads to the 964 elements that form the base set of detections. Being FPDW a monolithic detector, it is constrained to generate detections which lie completely inside the image boundary. This naturally generates a detection set without persons truncated by the image boundary, facilitating the re-identification (RE-ID).

Re-identification. We use the state-of-the-art RE-ID algorithm from (Figueira et al., 2013). Given a BB provided by the GT or the PD module, the algorithm first detects body parts using Pictorial Structures (Andriluka et al., 2009) (see Figure 3). The algorithm uses a body model consisting of head, torso, two thighs and two shins during the detection phase. It then merges the areas corresponding to both thighs and both shins, respectively. Subsequently, the algorithm extracts color histograms and texture histograms from the distinct image regions. Finally it performs the classification based on the extracted features with the Multi-View algorithm (Quang et al., 2013).

The necessary GT for the RE-ID task is obtained by processing the original GT and the detections generated by the PD module. Each detection is associated with the label of a person or the special label for the FP class. The assignment is done associating each detection with the label of the GT BB that has the most overlap with it. The Pascal VOC criterion (Everingham et al., 2010) is used to determine FP's: when the intersection between a detection BB and the corresponding BB from the original GT is smaller than half the union of the two, the detection is marked as a FP.

We use the Cumulative Matching Characteristic curve (CMC), the *de facto* standard, for evaluating the performance of RE-ID algorithms. The CMC shows how often, on average, the correct person ID is included in the best K matches against the training set for each test image. The overall performance is measured by the nAUC – the normalized Area Under the CMC curve.

4.1 Experiments

We evaluate the performance of different configurations of the RE-ID system as detailed in Table 1. Initially, we assess the performance of the RE-ID module in the conditions that are common in the RE-ID literature: using manually labelled GT BB's as test set. This is done to establish a term of comparison. Then we evaluate the naive integration of the PD and RE-ID modules and we contrast that with the integrated system enriched with the introduction of the FP class in the RE-ID module. Eventually, we assess the effectiveness of two operating modes for the Occlusion Filter, on the integrated system using the FP class.

In exp. (A) we perform RE-ID on the 416 fully visible BB's provided in the GT, consistently with the *modus operandi* of the state of the art. This means that the RE-ID module works with unoccluded persons and BB's that are correctly centred and sized. This provides a meaningful term of comparison when moving towards a fully automatic RE-ID system (see

Table 1: **GT** indicates the use of Ground Truth in an experiment, namely the hand-labelled Bounding Boxes. In **Occlusion Filter**, ALL indicates the "RejectAll" mode, and FAR the "RejectFarther" mode. NA indicates that the use of the Occlusion Filter or the FP class is not applicable to experiments with GT data. **Rank1** corresponds to the first point of the CMC curve and indicates how often the correct person ID is the best match against the training set. **nAUC** corresponds to the area under the CMC curve. The total number of detections and the amount of corresponding false positives which are passed to the RE-ID module are listed under **Detections** and **False Positives**, respectively.

Exp.	GT	Occlusion Filter	FP class	Rank1 (%)	nAUC (%)	Detections	False Positives
A	1	NA	NA	42.02	90.21	416	0
B	0	OFF	OFF	26.45	72.68	964	155
C	0	OFF	ON	31.33	87.60	964	155
D	0	ON ALL	ON	39.61	88.01	563	88
E	0	ON FAR 30%	ON	34.19	89.14	854	119

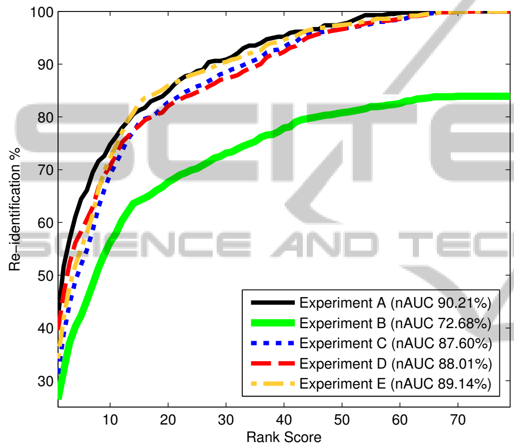


Figure 6: Cumulative Matching Characteristic curves comparing the performance of various configurations of the integrated Re-Identification system, for details see Table 1. Working points which lie comparatively higher or more to the left of the plot correspond to better performances.

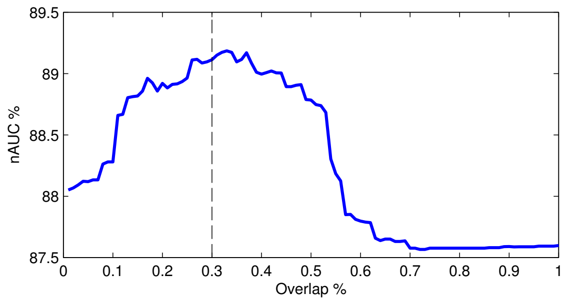


Figure 7: Normalized area under the Cumulative Matching Characteristic curve plotted against varying degrees of overlap for the Overlap Filter in 'RejectFarther' mode. Based on this evidence we set this parameter for exp. (E) to 30%.

Fig. 6 for the CMC curves associated to this and to other experiments). It is to be noted that this method of operation is not applicable in a real-world scenario, since it requires manual annotation of every person in the video sequence.

In exp. (B) we analyse the performance of the system resulting from the naive integration of the PD and RE-ID modules. The 155 FP's generated by the detector are impossible for the RE-ID to classify correctly, yielding a CMC curve that does not reach 100%. Applying the CMC evaluation scheme to this case leads to an unfair comparison with the curve from exp. (A). We provide this result to underline this point.

Exp. (C) shows that using the FP class in the RE-ID module allows for a meaningful comparison with the classic evaluation approach – exp. (A). The RE-ID module is able to classify a fraction of the FP's as such, thus achieving a performance comparable to that of exp. (A).

Rejecting all the detections affected by mutual overlap – exp. (D) – leads to a drastic reduction of the number of detections presented in input to the RE-ID module. This rejection mode eliminates 42% of the initial detections, including many depicting unoccluded pedestrians. This is due to this filtering method not taking advantage of perspective information. In Figure 4, for instance, one of the persons is fully visible in spite of the overlap between the bounding boxes. Because the CMC evaluation is concerned solely with the quality of the re-identification of the test examples it is provided with, it does not penalise the drastic reduction in detections. A practical application of RE-ID, though, would certainly be adversely affected by the Missed Detections.

Finally, in exp. (E) we run the Overlap Filter in the the "RejectFarther" mode. Figure 7 shows the impact of varying the filter's overlap threshold. We empirically choose 30% as the best value of this parameter, and illustrate again in Figure 6 the corresponding CMC curve. The "RejectFarther" mode proves to be better than the "RejectAll" both in terms of classification performance (showing a 1% increase in nAUC) and of the number of Missed Detections (rejecting only 12% of the initial detections).

5 CONCLUSIONS

In this work we studied how to combine person detectors and re-identification algorithms towards the development of fully automated surveillance systems. Whereas current work on re-identification focuses on matching "clean data", we analyse the effect of the unavoidable "noise" produced by automatic detections methods and devise ways to deal with it. In particular, a correct treatment of false positives and occlusions can recover performance levels that are lost if the modules are naively combined. This requires some additional characterisation of the acquisition process, in terms of collecting false positives in the environment where the system is deployed and performing a geometrical analysis to define the overlap filter criteria. This effort is largely compensated by the obtained improvements. In future work we aim at extending our methodology and evaluation criteria to drop the closed-space assumption and address the effect of missed detections in the quality and usability of fully automatic re-identification systems.

ACKNOWLEDGEMENTS

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] and the project High Definition Analytics (HDA), QREN - I&D em Co-Promoção 13750.

REFERENCES

Andriluka, M., Roth, S., and Schiele, B. (2009). Pictorial Structures Revisited: People Detection and Articulated Pose Estimation. *CVPR*.

Bak, S., Corvée, E., Brémond, F., and Thonnat, M. (2012). Boosted human re-identification using Riemannian manifolds. *ImaVis*.

Cheng, D. S., Cristani, M., Stoppa, M., Bazzani, L., and Murino, V. (2011). Custom pictorial structures for re-identification. In *BMVC*.

Corvee, E., Bak, S., and Bremond, F. (2012). People detection and re-identification for multi surveillance cameras. *VISAPP*.

Dikmen, M., Akbas, E., Huang, T., and Ahuja, N. (2011). Pedestrian recognition with a learned metric. In *ACCV*.

Dollár, P., Belongie, S., and Perona, P. (2010). The Fastest Pedestrian Detector in the West. *BMVC*.

Ess, A., Leibe, B., and Van Gool, L. (2007). Depth and Appearance for Mobile Scene Analysis. *ICCV*.

Everingham, M., Van Gool, L., Williams, C., Winn, J., and Zisserman, A. (2010). The Pascal visual object classes (VOC) challenge. *IJCV*.

Farenzena, M., Bazzani, L., Perina, A., Murino, V., and Cristani, M. (2010). Person re-identification by symmetry-driven accumulation of local features. In *CVPR*.

Figueira, D., Bazzani, L., Minh, H. Q., Cristani, M., Bernardino, A., and Murino, V. (2013). Semi-supervised multi-feature learning for person re-identification. *AVSS*.

Gheissari, N., Sebastian, T., and Hartley, R. (2006). Person reidentification using spatiotemporal appearance. In *CVPR*.

Girshick, R., Felzenszwalb, P., and McAllester, D. (2011). Object detection with grammar models. *PAMI*.

Gray, D. and Tao, H. (2008). Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*.

Hamdoun, O., Moutarde, F., Stanciulescu, B., and Steux., B. (2008). Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *ICDSC*.

Harandi, M. T., Sanderson, C., Wiliem, A., and Lovell, B. C. (2012). Kernel analysis over Riemannian manifolds for visual recognition of actions, pedestrians and textures. In *WACV*.

Hirzer, M., Roth, P., Kostinger, M., and Bischof, H. (2012). Relaxed pairwise learned metric for person re-identification. In *ECCV*.

Jungling, K., Bodensteiner, C., and Arens, M. (2011). Person re-identification in multi-camera networks. In *CVPRW*.

Li, W. and Wang, X. (2013). Locally aligned feature transforms across views. In *CVPR*.

Li, W., Zhao, R., and Wang, X. (2012). Human reidentification with transferred metric learning. In *ACCV*.

Liu, C., Gong, S., Loy, C., and Lin, X. (2012a). Person re-identification: What features are important? In *ECCV*.

Liu, C., Wang, G., Lin, X., and Li, L. (2012b). Person re-identification by spatial pyramid color representation and local region matching. *IEICE*.

Ma, B., Su, Y., and Jurie., F. (2012). Bicov: a novel image representation for person re-identification and face verification. In *BMVC*.

Mignon, A. and Jurie, F. (2012). Pcca: A new approach for distance learning from sparse pairwise constraints. In *CVPR*.

Mogelmose, A., Bahnsen, C., and Moeslung, T. B. (2013a). Tri-modal person re-identification with RGB, depth and thermal features. In *IEEE WPBVS*.

Mogelmose, A., Moeslund, T. B., and Nasrollahi, K. (2013b). Multimodal person re-identification using RGB-D sensors and a transient identification database. In *IWBF*.

Pedagadi, S., Orwell, J., Velastin, S., and Boghossian, B. (2013). Local fisher discriminant analysis for pedestrian re-identification. In *CVPR*.

Prosser, B., Zheng, W., Gong, S., Xiang, T., and Mary., Q. (2010). Person re-identification by support vector ranking. In *BMCV*.

Quang, M. H., Bazzani, L., and Murino, V. (2013). A unifying framework for vector-valued manifold regularization and multi-view learning. In *ICML*.

Taiana, M., Nascimento, J., and Bernardino, A. (2013). An improved labelling for the INRIA person data set for pedestrian detection. *IbPRIA*.

Wang, X., Doretto, G., Sebastian, T., Rittscher, J., and Tu, P. (2007). Shape and appearance context modeling. In *ICCV*.

Wu, Y., Mukunoki, M., Funatomi, T., Minoh, M., and Lao, S. (2011). Optimizing mean reciprocal rank for person re-identification. In *AVSS*.

Zheng, W., Gong, S., and Xiang, T. (2009). Associating groups of people. In *BMVC*.

Zheng, W., Gong, S., and Xiang, T. (2011). Person re-identification by probabilistic relative distance comparison. In *CVPR*.