Comparative Analysis of PRID Algorithms Based on Results Ambiguity Evaluation

V. Renò[1(✉)], A. Cardellicchio[2], T. Politi[2], C. Guaragnella[2], and T. D'Orazio[1]

[1] Institute of Intelligent Systems for Automation,
Italian National Research Council, via Amendola 122 D/O, 70126 Bari, Italy
`reno@ba.issia.cnr.it`
[2] Dipartimento di Ingegneria Elettrica e dellInformazione,
Politecnico di Bari, via Orabona 4, 70126 Bari, Italy

Abstract. The re-identification of a subject among different cameras (namely Person Re-Identification or *PRID*) is a task that implicitly defines ambiguities. Two individuals dressed in a similar manner or with a comparable body shape are likely to be misclassified by a computer vision system, especially when only poor quality images are available (i.e. the case of many surveillance systems). For this reason we introduce a method to find, exploit and classify ambiguities among the results of *PRID* algorithms. This approach is useful to analyze the results of a classical PRID pipeline on a specific dataset evaluating its effectiveness in re-identification terms with respect to the ambiguity rate (AR) value. Cumulative Matching Characteristic curves (CMC) can be consequently split according to the AR, using the proposed method to evaluate the performance of an algorithm in low, medium or high ambiguity cases. Experiments on state-of-art algorithms demonstrate that ambiguity-wise separation of results is an helpful tool in order to better understand the effective behaviour of a PRID approach.

1 Introduction

Person re-identification (PRID) is a crucial task in modern video surveillance systems, and concerns the retrieval of the same individual given several views acquired by a set of non-overlapping cameras. PRID is strictly related to a number of other video surveillance topics, like cross-camera tracking, event analysis, abandoned object retrieval, and so on; however, it is an extremely challenging task, and has recently drawn a lot of focus by researchers with different fields of expertise.

First of all, each camera in a surveillance system has specific hardware properties that, together with varying lighting conditions, introduce slight variations in the captured frames which, as a consequence, have to be conducted to a common baseline using proper image processing techniques. Furthermore, pose variations of the subject, along with occlusion phenomena, have to be taken into account, as they can negatively impact PRID performances hiding discriminating features that could be otherwise exploited.

Another challenging issue is related to the method used to evaluate the performance of PRID algorithms. Normally, PRID techniques are tested against one or more datasets, each one with specific characteristics and challenges. In a pre-processing step, the considered dataset is split into a gallery set, which contains exactly one view per individual, and a probe set, which contains one or more views per subject. The algorithm compares each instance of the gallery set against a set of views taken by the probe set, searching for the best possible match; therefore, PRID can be seen as a multiclass classification problem, where each subset of views related to a certain individual represents a specific class. The task of determining meaningful features – and the proper classifier that should be used – is non-trivial; furthermore, there is not an *universal* dataset (i.e. a dataset that can be used to test methodologies against every specific issue) and, as a consequence, an algorithm which gives good results on a certain dataset may obtain mediocre performances on another dataset. Traditional PRID approaches deal with these problems using a recurring scheme to which we will refer to as PRID pipeline [1]. The first step in the PRID pipeline is image segmentation, where significant information are extracted using proper pre-processing techniques (i.e. background subtraction ([2–6]), human detection ([7,8]) and shadow suppression ([9])). In the second step, a discriminating signature is computed for each view, starting from robust features which can be related to appearance (i.e. color, texture, or shape [10–13]) or to other characteristics like gait [14]. In the third and last step, signatures extracted in the previous step are compared to find the most similar image pairs. Classic matching methods exploited fixed metrics, like Euclidean distance or Bhattacharyya coefficient; modern methods employs more sophisticated approaches, as distance metric learning ([15–17]) or machine learning ([18,19]). Recently, the whole pipeline has been replaced by deep learning architectures ([20–22]), which automatically extract discriminating features at different levels of abstraction, combining them into a meaningful signature, thus giving a significant boost in terms of performances.

Even the most sophisticated state of the art approaches still rely on a basic assumption: results comparison is carried out using an agnostic method based on the analysis of Cumulative Matching Characteristics (CMC) curves, as reported in Fig. 1. These curves describe re-identification results in terms of ranking, that represents the number of iterations after which the PRID algorithm is able to output the correct match. Specifically, Rank-1 represents correctly matched subjects, Rank-2 shows how many individuals are being re-identified after one iteration, and so on. However, this approach only allows us to understand the recognition percentage at a specific rank, without adding any specific information on how results have been produced. Hence, CMC curves give quantitative results, without taking into account neither the intrinsic difficulties of a given dataset, nor the qualitative meaning of the achieved results. For example, given a certain rank, it is not possible to understand if the result has been achieved comparing pair of samples which show low, medium or high ambiguity properties [24]. In case of low ambiguity, the correct match should be ideally returned
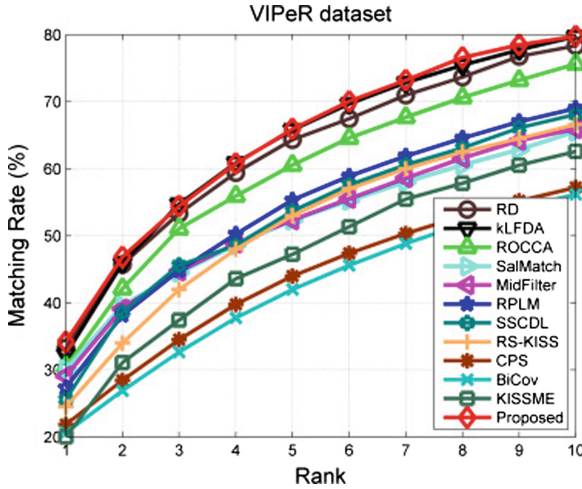
Fig. 1. Example of overall CMC curve taken from [23].

as a Rank-1, while worst results should be expected when the algorithm faces ambiguous observations. Normally, the most significant results should be return within Rank-5.

Furthermore, a graphical overview of a certain number of examples taken by VIPeR is depicted in Fig. 2. In this example, some images are categorized in easy, ambiguous or difficult cases. It is immediate to notice that the first are cases in which the human vision system is able to match correspondences easily, for example the textured sweater of the first subject, or red and yellow sweatshirts in the other images. Ambiguous cases are challenging situations even for an expert human operator because each subject looks like many others, especially when he is wearing dark clothes. Finally, difficult cases are always included in datasets and represent a cluster of images in which different light conditions and subject orientations make the correct association almost impossible. In these cases, a correct association would be probably due to fortuity rather than the effective recognition of features by the PRID algorithm.

With this work, we further explore the concept of Ambiguity Rate (AR) introduced in [24] – i.e. an index that compares the results given by a PRID algorithm on a specific dataset – evaluating the performances of state-of-the art PRID algorithms in predetermined ambiguity ranges. The rest of the paper is organized as follows. In Sect. 2, we will give an explanation of our method, highlighting the algorithm used to extract the AR. In Sect. 3, we will compare the results of three PRID approaches using AR, while in Sect. 4 conclusions and a perspective on future works is given.

Fig. 2. Example of subjects from VIPeR dataset. The first row represents some query images, while the second row contains the corresponding ground truth. In this example pictures have been manually clustered in easy, ambiguous or difficult to recognize and are respectively highlighted with a blue, black or red contour rectangle. (Color figure online)

2 Methodology

2.1 Algorithm Description

The proposed approach is related to the one presented in [24] and can be summarized as an enhancement of the PRID pipeline in terms of results' analysis.
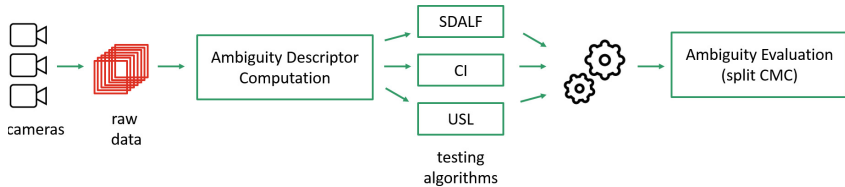

Fig. 3. Algorithm high level block diagram.

Looking at Fig. 3 it is immediate to notice that the testing algorithms only represent the central part of the whole approach. In fact, raw data (i.e. images coming from a video surveillance system or a dataset) are first of all pre-processed in order to compute the ambiguity descriptor while the ambiguity evaluation is done as the last step, when results in terms of iterations and ranks are available for each algorithm.

2.2 Pre-processing Step

An interesting aspect of this methodology is that the framework does not impose strict constraints in the definition of the ambiguity descriptor ad. This entity

numerically describes the particular scene or the specific image patch that is going to be analyzed by a PRID algorithm using different features that can be chosen according to the phenomenon that needs to be investigated. For example, a dataset rich of people that wear textured clothes will probably be described in terms of textural features, while general purpose datasets will rely on color based descriptions. Assuming that an expert video surveillance operator evaluates the output of a semi-automatic system by observing recurrent colors in the images, ambiguities in this paper are defined in terms of color changes. Therefore, for each frame of the input dataset, we define the ambiguity descriptor as an array of color values $ad = [h_1, h_2, \ldots, h_n]^T$ where n is the number of horizontal stripes used to divide the image and h_k represents the modal value of the Hue coordinate of the k-th stripe.

2.3 Post-processing Steps

Ambiguity can be evaluated after executing testing algorithms on the chosen dataset. These methods basically associate a similarity score to each image pair that combines one sample from the gallery set with all the samples from the probe set. For example, the results obtained for a query image q_i can be represented as an associative array

$$q_i \Longleftrightarrow [r_1, r_2, \ldots, r_M, \ldots, r_P] \tag{1}$$

where P is the number of images in the probe set and M is a threshold used to consider only the best results. $[r_1, r_2, \ldots, r_M]$ can be therefore represented in terms of the ambiguity descriptor chosen in the pre-processing step, defining the Ambiguity Descriptor Matrix for the query image q_i

$$
\begin{aligned}
ADM_{q_i} &= [ad_{r_1}, ad_{r_2}, \ldots, ad_{r_M}] \\
&= \begin{pmatrix} h_{1r_1} & h_{1r_2} & \ldots & h_{1r_M} \\ h_{2r_1} & h_{2r_2} & \ldots & h_{2r_M} \\ \vdots & \vdots & \vdots & \vdots \\ h_{nr_1} & h_{nr_2} & \ldots & h_{nr_M} \end{pmatrix} = \begin{pmatrix} hS_1^T \\ hS_2^T \\ \vdots \\ hS_n^T \end{pmatrix}
\end{aligned} \tag{2}
$$

where hS_j^T is the array in which the modal values of the best M frames of the j-th stripe are stored. These rows are employed to compute percentage deviations of color features without losing spacial information by applying the following formula:

$$\%_{q_i} = \begin{pmatrix} \dfrac{\max(hS_1^T) - \min(hS_1^T)}{256} \\ \vdots \\ \dfrac{\max(hS_n^T) - \min(hS_n^T)}{256} \end{pmatrix} \tag{3}$$

Finally, the AR value for q_i is defined starting from the average value of the percentage deviations

$$AR_{q_i} = 1 - \frac{1}{n} \sum_{s=1}^{n} \%_{q_i}(s) \tag{4}$$

so that low variations of color percentage displacements produce high ambiguity rate. The alternation of colors in the results, instead, are related to low AR values. It is worth noticing that the role of M is essential for the ambiguity to be effectively computed because useful information can be extracted only within the best ranks. If all the images from the probe set were used to compute the ambiguity rate, AR would be exactly the same for each query image.

Finally, CMC curves can be split according to the ambiguity rate simply setting thresholds and filtering the results. In this paper we define three different ambiguity ranges according to the following fuzzification rule:

$$
\begin{aligned}
R_1 &\to 0 \leq AR \leq 0.4 \\
R_2 &\to 0.4 < AR \leq 0.8 \\
R_3 &\to 0.8 < AR \leq 1
\end{aligned}
\tag{5}
$$

This way CMC curves can be drawn considering multiple contributions: the first for low ambiguity rates, the second for medium ones and the third for high ambiguity results.

3   Experiments and Results

Ambiguity evaluation as described in the previous sections has been performed on three person re-identification algorithms known in literature:

– **Symmetry-Driven Accumulation of Local Features (SDALF)** [10], that basically exploits color (Maximally Stable Color Regions and Weighted Color Histograms) and texture (Recurrent High-Structured Patches) features around pedestrian symmetry axes for extracting image signatures;
– **Color Invariants for PRID (CI)** [25], that exploits relationships between different color patches extracted from each pedestrian image (usually two: one for the upper part and the other for the lower part);
– **Unsupervised Salience Learning for PRID (USL)** [26], that exploits salient features (unique and discriminative) to characterize each person. Both color histograms and SIFT features are used to extract signatures that are subsequently processed by a classifier (e.g. SVM or KNN).

All the algorithms have been tested on the well known VIPeR dataset [27], that contains 632 images taken from non overlapping cameras with arbitrary viewpoints. Images belonging to VIPeR have been taken under varying illumination conditions and each one is scaled to 128×48 pixels. The approach presented in this paper is mainly focused on the interpretation of results on split CMC curves, according to the fuzzification rule presented in the previous section. Different curves for different ambiguity rate values help in better understanding the algorithm capabilities and interpret if it is producing ambiguous results or not. For this reason, the experiments presented in this section will exploit the AR value to understand how a specific algorithm is working given a specific ambiguity range.

Table 1. Dataset separation for different values of the ambiguity rate.

Ambiguity rate	SDALF		CI		USL	
LOW	16 img	5.06%	18 img	5.70%	18 img	5.70%
MEDIUM	298 img	94.30%	287 img	90.82%	279 img	88.29%
HIGH	2 img	0.63%	11 img	3.48%	19 img	6.01%

First, the ambiguity descriptor is computed as described in Sect. 2.2 for each image of the collection using 6 horizontal stripes. Then, the rest of the PRID pipeline is executed for the chosen algorithms and finally the results are processed in order to compute the ambiguity rate. In order to obtain a visual comparison of the least ambiguous result and the most ambiguous one, examples of boxplot enriched by the corresponding frames are provided in Figs. 4 and 5. Each box refers to one of the stripes used to divide the images, as noticeable in the figure, so it is representative of ADM described in Eq. 2. A boxplot with large boxes will refer to a non ambiguous response, that should basically imply that the algorithm is operating in an *easy condition*, so the correct response should be given at the first rank. On the contrary, small boxes are related to ambiguous responses that are likely to be mistaken. In this situation, a good PRID algorithm should return the correct answer within the first ranks, but not always at rank 1. Looking at Fig. 4, the first thing to point out is that the only algorithm able to re-identify the query image is CI at rank 3. SDALF is not producing the correct answer in five ranks. Images at ranks $1, 2, 3$ depict people with beige trousers, but the upper part of the query image is not being considered by this approach. The case of USL suggests that there is a consistent amount of people that are likely to be misclassified due to extremely similar clothing. The analysis of Fig. 5 shows that images taken within the best results actually are not so ambiguous. The first 5 returned values are different one from the other: different colours of the shirt/dress (red, black, green, orange and gray) and different colours of the trousers/skirt (pink, black, red, orange). In this situation, the only algorithm that is not answering correctly is CI, while SDALF and USL achieve a rank 1 result. Both the results of CI (for the minimum AR case) and USL (for the maximum AR case) show how the features used by the algorithms can not isolate easy recognizable situations for a human eye. This is probably due to the representation of the colors in different visual systems: the human one and the digital one. For the first, peaks on different color tones can be immediately distinguishable, while in a digital color space the same peaks can generate values that are likely to be classified as similar colors even if they are different. This suggests us to investigate a methodology to quantify the global ambiguity of a dataset and associate an ambiguity level to each query image (e.g. easy, medium or difficult), as will be discussed in the future works section.

Figure 6 shows the ambiguity rate histogram for each response of the three algorithms. The background of the plot helps in the visualization of the three ranges: it is immediate to notice that a small number of responses has a
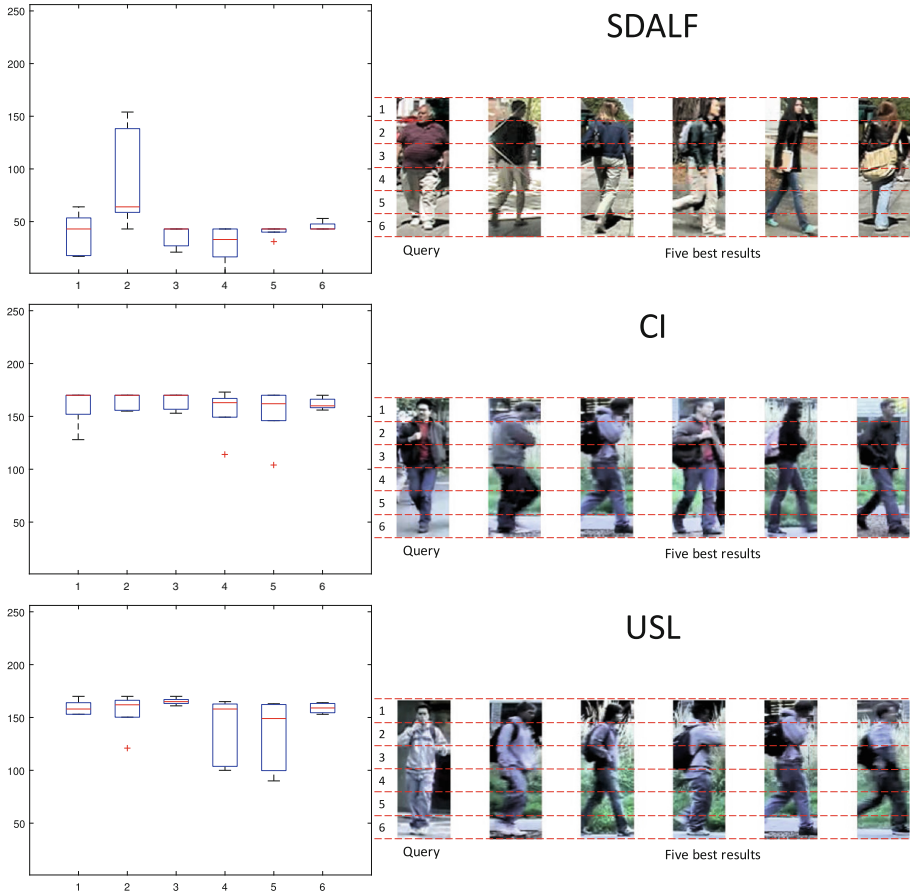
Maximum ambiguity rate boxplot

Fig. 4. Boxplots that represent the result with the highest ambiguity rate for each algorithm. The plot is enriched with the visual information about both query image and the first five results returned by the specific algorithm. The division stripes are reported on the x-axis, while mode values of hue coordinates are plotted on the y-axis. Small boxes are referred to high ambiguity and big boxes to low ambiguity.

corresponding low ambiguity rate (< 0.4) or a high one (> 0.8), according to the fuzzification rule presented beforehand.

Looking at Table 1, all algorithms are isolating a small percentage of the images in the tails of the distribution, namely the 5% of the results of the algorithms has low ambiguity. The behaviour for high ambiguity rates is different: only 2 images actually fall into this category for SDALF, 11 for CI and 19 for USL. This means that the algorithms tend to avoid extremely ambiguous responses. A medium level ambiguity is produced most of the time, as almost
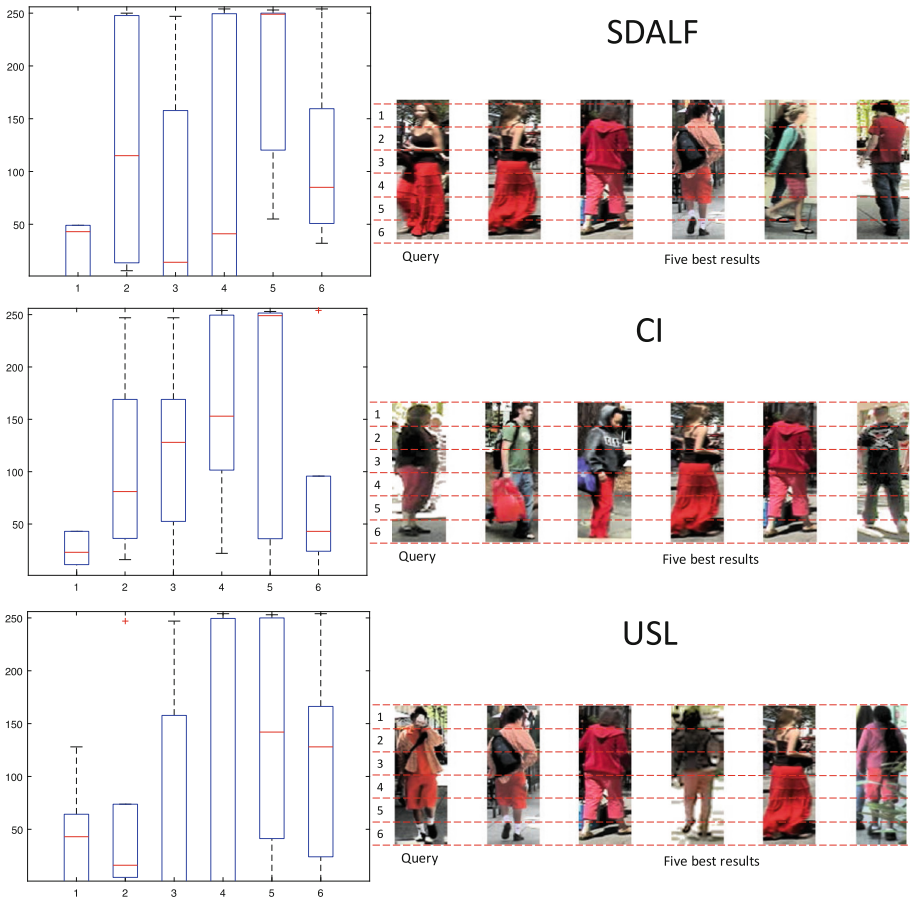
Minimum ambiguity rate boxplot

Fig. 5. Boxplots that represent the result with the lowest ambiguity rate for each algorithm. The plot is enriched with the visual information about both query image and the first five results returned by the specific algorithm. The division stripes are reported on the x-axis, while mode values of hue coordinates are plotted on the y-axis. Small boxes are referred to high ambiguity and big boxes to low ambiguity. (Color figure online)

90% have an AR value between 0.4 and 0.8. The corresponding CMC curves for LOW, $MEDIUM$ and $HIGH$ ambiguity rate values are reported in Fig. 7 and are called *split CMC*. For each curve, the x axis reports the first 100 ranks and the y axis shows the percentage of images that have been recognized at the specific rank. Due to the cumulative nature of the curve, if there is a step, it means that there are no matches at the corresponding rank. An example of this behaviour is shown in Fig. 7 (d), where there is a big step that starts
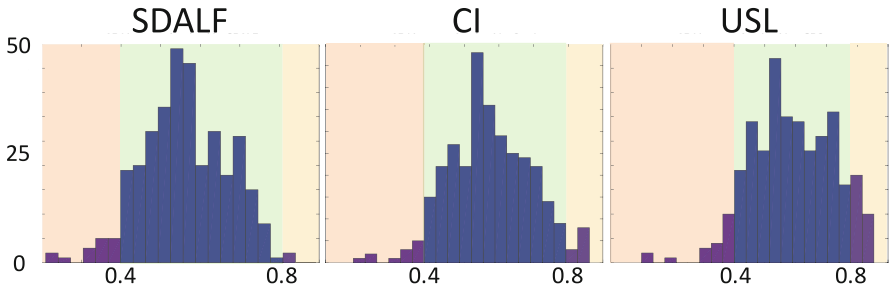
Ambiguity Rate Histograms

Fig. 6. Ambiguity rate histograms on VIPeR results. The x-axis represents the ambiguity rate while on the y-axis occurrences are counted. Low, medium and high ambiguity rate ranges are respectively orange, green and yellow highlighted. (Color figure online)
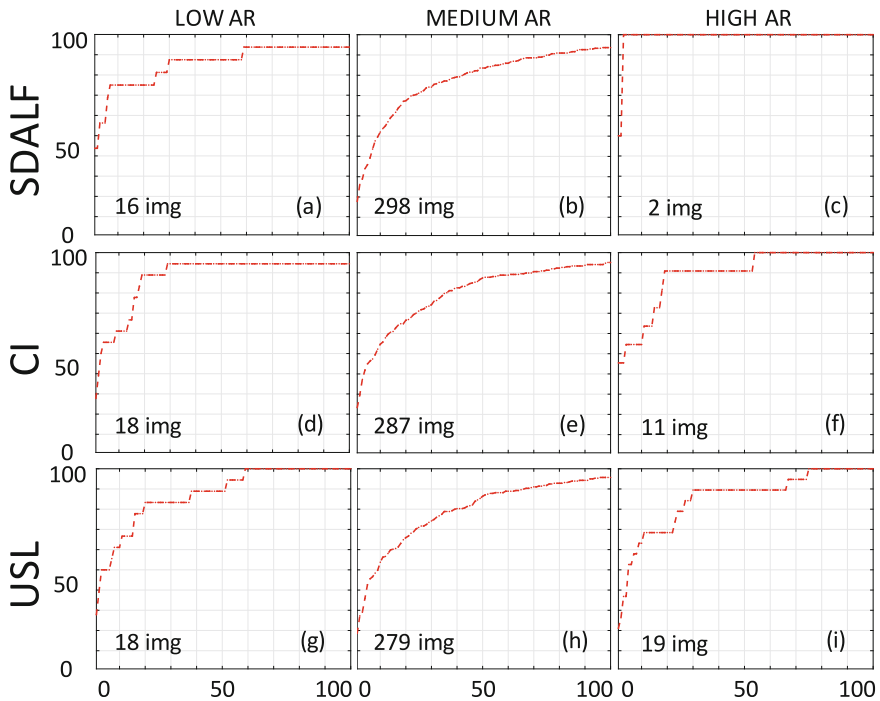
Split CMC curves

Fig. 7. Split cumulative matching characteristic curves obtained for the three considered algorithms.

approximatively around rank 35. The easiest operating condition for an algorithm, where the expected result would be a really high percentage at rank 1, is the *LOW* AR. Here, the best algorithm in our experiments is SDALF because it achieves about 50% of results at rank 1 and about 80% of results within the first ranks. The other approaches obtain a similar result with more iterations. The CMCs in Fig. 7 (b), (e) and (h) are similar to the ones already known in literature because they are representative of about 90% of the dataset. A final remark should be pointed for *HIGH* ambiguity rates. SDALF seems to be the best algorithm in this comparison (with 50% rank 1 responses and 100% rank 2), but the cardinality of the *HIGH* AR set is only 2. This means that the two images are immediately recognized by the algorithm, even if it is working in a challenging situation. Both CI and USL show comparable results in the middle of curves (f) and (i), where there is a step for a recognition percentage of about 90%. CI reports a good starting point, as its rank 1 accuracy is about 50%, while on the other hand USL is able to gain its performances within the first ranks, passing from 20% to 70% in a couple of iterations. Independently from a particular experiment, a generic algorithm should be able to increase the number of images that lie in the tail of its ambiguity distribution. When dealing with *LOW* AR values, the recognition percentage at rank 1 should be the highest, while the correct response can be expected within the first ranks for *HIGH* ambiguity queries.

4   Conclusion

In this paper, ambiguities have been exploited in order to evaluate the accuracy of a re-identification algorithm splitting well known CMC curves. The methodology basically defines an ambiguity descriptor and relies on it to compute the AR of each query performed by an algorithm on a specific dataset, actually enriching the state-of-art PRID pipeline. The definition of ambiguity evaluated in this paper can be seen as a *relative* one, because it depends on the results that the algorithm achieves on each query, as stated in Eq. 2. The AR histogram (Fig. 6) graphically explains the ambiguity distribution among the images of a specific dataset, while split CMC curves can be studied separately (ambiguous vs. non ambiguous situations), enabling us to measure the performance of different algorithms on the same dataset. However, the work presented in this paper is the first step in the exploitation of ambiguities in order to understand the capabilities of a re-identification approach. Even if *relative* ambiguity modelling is certainly useful to understand the operative conditions in which an algorithm is working, the results shown in Sect. 3 inspire future research in the direction of an *absolute* ambiguity definition. This way, each image of a dataset will be classified as *easy* (e.g. the only orange dressed man in a crowd of dark clothed subjects) or *difficult* (e.g. a black dressed man in a crowd of dark clothed people). Finally, exploiting both *relative* and *absolute* ambiguities, a generic rank of a CMC will be promoted or penalized starting from the assumption that easy cases should not be misclassified, while higher ranks can be tolerated for hard queries.

References

1. Cardellicchio, A., D'Orazio, T., Politi, T., Renò, V.: An human perceptive model for person re-identification. In: VISAPP-International Conference on Computer Vision Theory and Applications-2015 (2015)
2. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2. IEEE (1999)
3. Zivkovic, Z.: Improved adaptive Gaussian mixture model for background subtraction. In: 17th International Conference on Pattern Recognition, ICPR 2004, vol. 2, pp. 28–31 (2004)
4. Jojic, N., Perina, A., Cristani, M., Murino, V., Frey, B.: Stel component analysis: Modeling spatial correlations in image class structure. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 2044–2051. IEEE (2009)
5. Renò, V., Marani, R., D'Orazio, T., Stella, E., Nitti, M.: An adaptive parallel background model for high-throughput video applications and smart cameras embedding. In: Proceedings of the International Conference on Distributed Smart Cameras, ICDSC 2014, pp. 30:1–30:6. ACM, New York (2014)
6. Spagnolo, P., Leo, M., D'Orazio, T., Distante, A.: Robust moving objects segmentation by background subtraction. In: The International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS) (2004)
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
8. Corvee, E., Bak, S., Bremond, F., et al.: People detection and re-identification for multi surveillance cameras. In: International Conference on Computer Vision Theory and Applications, VISAPP-2012 (2012)
9. Lu, J., Zhang, E.: Gait recognition for human identification based on ICA and fuzzy SVM through multiple views fusion. Pattern Recognit. Lett. **28**, 2401–2411 (2007)
10. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2360–2367. IEEE (2010)
11. Gheissari, N., Sebastian, T.B., Hartley, R.: Person re-identification using spatiotemporal appearance. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1528–1535. IEEE (2006)
12. Roy, A., Sural, S., Mukherjee, J.: A hierarchical method combining gait and phase of motion with spatiotemporal model for person re-identification. Pattern Recognit. Lett. **33**, 1891–1901 (2012)
13. D'Orazio, T., Guaragnella, C.: A graph-based signature generation for people re-identification in a multi-camera surveillance system. In: VISAPP, vol. 1, pp. 414–417 (2012)
14. Bauml, M., Stiefelhagen, R.: Evaluation of local features for person re-identification in image sequences. In: 2011 8th IEEE International Conference on Advanced Video and Signal-Based Surveillence (AVSS), pp. 291–296. IEEE (2011)
15. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 649–656. IEEE (2011)

16. Hirzer, M., Roth, P.M., Köstinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7577, pp. 780–793. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33783-3_56

17. Köstinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2288–2295. IEEE (2012)

18. Prosser, B., Zheng, W.S., Gong, S., Xiang, T., Mary, Q.: Person re-identification by support vector ranking. In: BMVC, vol. 2, p. 6 (2010)

19. Layne, R., Hospedales, T.M., Gong, S., Mary, Q.: Person re-identification by attributes. In: BMVC, vol. 2, p. 8 (2012)

20. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 152–159 (2014)

21. Ahmed, E., Jones, M., Marks, T.K.: An improved deep learning architecture for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3908–3916 (2015)

22. Ding, S., Lin, L., Wang, G., Chao, H.: Deep feature learning with relative distance comparison for person re-identification. Pattern Recogn. **48**, 2993–3003 (2015)

23. An, L., Chen, X., Yang, S.: Person re-identification via hypergraph-based matching. Neurocomputing **182**, 247–254 (2016)

24. Renò, V., Cardellicchio, A., Politi, T., Guaragnella, C., D'Orazio, T.: Exploiting ambiguities in the analysis of cumulative matching curves for person re-identification. In: ICPRAM 2016 - Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods, pp. 484–494 (2016)

25. Kviatkovsky, I., Adam, A., Rivlin, E.: Color invariants for person reidentification. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1622–1634 (2013)

26. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3586–3593 (2013)

27. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: Proceedings of IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), vol. 3. Citeseer (2007)