Open-set Person Re-identification

Shengcai Liao, Zhipeng Mo, Jianqing Zhu, Yang Hu, and Stan Z. Li

Center for Biometrics and Security Research
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing 100190, China
{scliao,jianqing.zhu,yhu,szli}@nlpr.ia.ac.cn

Abstract. Person re-identification is becoming a hot research for developing both machine learning algorithms and video surveillance applications. The task of person re-identification is to determine which person in a gallery has the same identity to a probe image. This task basically assumes that the subject of the probe image belongs to the gallery, that is, the gallery contains this person. However, in practical applications such as searching a suspect in a video, this assumption is usually not true. In this paper, we consider the open-set person re-identification problem, which includes two sub-tasks, detection and identification. The detection sub-task is to determine the presence of the probe subject in the gallery, and the identification sub-task is to determine which person in the gallery has the same identity as the accepted probe. We present a database collected from a video surveillance setting of 6 cameras, with 200 persons and 7,413 images segmented. Based on this database, we develop a benchmark protocol for evaluating the performance under the open-set person re-identification scenario. Several popular metric learning algorithms for person re-identification have been evaluated as baselines. From the baseline performance, we observe that the open-set person re-identification problem is still largely unresolved, thus further attention and effort is needed.

Keywords: person re-identification, open-set identification, database and benchmark

1   Introduction

Video surveillance is an important technique for both security and forensic applications. However, since the volume of surveillance videos have become larger and larger today, searching desired objects from massive surveillance videos turns out to be very difficult. Person re-identification is one of the useful technique to facilitate this head-scratching search. It is becoming a hot research for developing both machine learning algorithms and video surveillance applications. Many approaches have been proposed for person re-identification [1, 2], which greatly advance this field.

The task of person re-identification is to determine which person in a gallery has the same identity to a probe image. This task basically assumes that the

subject of the probe image belongs to the gallery, that is, the gallery contains this person. This scenario is called the closed-set identification [3]. Under this assumption, the evaluation of person re-identification algorithms is done by performing the top-k retrieval and determining whether the true match appears in the result. Furthermore, the Cumulative Matching Characteristic (CMC) curve is used to illustrate the identification accuracy versus the number of samples retrieved. Many databases have been created to test person re-identification algorithms under the closed-set scenario, as summarized in Table 1.

A typical application of person re-identification in forensic video analysis is to search a suspect in a video or a large set of videos. In this application, the probe image contains a suspect (e.g. from a shot of a surveillance video). The task is to determine whether this suspect appears in a certain video, and if it does, when and where this person appears. In this application, there is no assumption that the person in the probe image is present in the searching video. Instead, the system needs to answer whether the query person appears in the searching video, and if it does, which person in the video corresponds to the query person. Without the ability of determining the presence of a suspect in a video, one can also apply available person re-identification algorithms to retrieve the top-k candidates, and browse the results to determine whether the suspect is found. However, considering that there may be a large set of videos to be searched, it still requires large human labors to check top-k candidates from every video. Instead, if a strong evidence indicates that the searching video does not contain the suspect, the video can be automatically discarded to save time and labor.

The open-set person re-identification problem studied in this paper is oriented for the above forensic application of searching a suspect in a recorded surveillance video or a large set of videos (treated as the gallery). In this application, two requirements for person re-identification algorithms are generally needed: i) algorithm should be trained with an independent data set, because it is not practical to learn a classifier or learn person models every time a video comes; and ii) algorithm should be able to answer whether the probe person is contained in the video or not. This problem is similar as the open-set face identification problem proposed in [3–5]. Formally, open-set person re-identification includes two sub-tasks, detection and identification. The detection sub-task is to determine the presence of the probe subject in the gallery, and the identification sub-task is to determine which person in the gallery has the same identity as the accepted probe.

Accordingly, in this paper, we present a database from a video surveillance setting of 6 cameras, with 200 persons and 7,413 images segmented. Based on this database, we develop a benchmark protocol for evaluating the performance under the open-set person re-identification scenario. Basically, the proposed open-set person re-identification benchmark protocol involves four subsets, namely, training set, gallery set, genuine probe set, and impostor probe set, where test set = gallery set + probe set, and probe set = genuine probe set + impostor probe set (see Fig. 3 for an illustration). With this setting we introduce two kinds of open-set problems: i) persons of the test set are not known from the training

set; and ii) persons of the impostor probe set are not known from the gallery set. We require learning a general camera independent person re-identification model from the training set, so as to test the algorithm's generalization ability to unseen persons and unknown camera views.

Several popular metric learning algorithms for person re-identification have been evaluated as baselines in this paper. However, from the baseline performance, we observe that the open-set person re-identification problem is still largely unresolved. Therefore, further effort should be spent to tackle this problem. We make the collected database (called Open-set *P*erson Re-IDentification database, OPeRID v1.0), as well as the benchmark tool and the extracted features publicly available in a project website[1] to promote further development along this direction.

The remaining paper is organized as follows. In Section 2, we briefly review the related work. In Section 3, we describe the collection of the database. In Section 4, we introduce the benchmark protocol for the open-set person re-identification. We introduce several metric learning algorithms as the baseline methods in Section 5, and present the baseline performance in Section 6. Finally, we conclude the paper in Section 7.

2   Related Work

Recently, more and more benchmark datasets for person re-identification are available. These datasets contain images of persons with variations in resolutions, lightings, poses, occlusion, and background in the same camera view and across camera views and thus make the person re-identification task very challenging. We summarize some popular person re-identification datasets in Table 1. Among them, the VIPeR dataset [6] is one of the earliest single-shot datasets in this field, and it is the most widely used benchmark dataset so far. The ETHZ dataset [7, 8] contains images captured from a single moving cameras in a street scene, therefore the main challenge is occlusions rather than viewpoint variations. The i-LIDS MCTS dataset [9, 10] was collected from an airport arrival hall, with some persons having images from the same camera view, while others having images from different non-overlapping camera views. The CAVIAR4REID dataset [11] contains 1,220 images of 72 pedestrians with large variation in resolutions, but only 50 of the 72 pedestrians appear in two camera views. The PRID2011 dataset [12] contains 931 persons from two camera views, with 200 persons appearing in both camera views. The GRID dataset [13] was captured from 8 disjoint camera views in a underground station. It contains 250 pedestrian image pairs, with each pair containing two images of the same person from different camera views. Besides, there are 775 additional images that do not belong to the 250 persons, which can be used to enlarge the gallery. The 3DPES dataset [14] contains 200 persons from 8 different surveillance cameras. An advantage of this dataset is that some persons can be seen from 3 different camera views. The CUHK02

---

[1] http://www.cbsr.ia.ac.cn/users/scliao/projects/operidv1/

dataset [15] contains 1,816 persons captured from 5 pairs of cameras. It has the largest number of persons so far. More details and characteristics of the above mentioned datasets are listed in Table 1.

Table 1. Some popular datasets used for person re-identification

Dataset	#Cams	#Person	#Imgs	#Views	Scenario
VIPeR [6]	2	632	1,264	2	Outdoor
i-LIDS [9, 10]	5	119	476	2	Indoor, Airport
ETHZ [7, 8]	3	146	8,555	2	Outdoor,City Street
CAVIAR4REID [11]	2	72	1,220	2	Indoor, Shopping Center
3DPeS [14]	8	200	1,012	3	Outdoor, Campus
PRID2011 [12]	2	200	1,134	2	Outdoor
CUHK02 [15]	5 pairs	1,816	7,264	2	Outdoor, Campus
GRID [13]	8	250	1,275	2	Indoor, Underground

Note: Column three for PRID2011 and GRID contains only persons appearing in both views. #Views denotes the maximum number of camera views for the same person.

The emergence of all these benchmark datasets have promoted the development of person re-identification to a large degree. However, they also have certain drawbacks that need to be addressed [2]. Here, we concern about the number of persons, samples, and the number of camera views that a person has. Most existing databases do not have enough data to support the four-subset division for open-set person re-identification as aforementioned. In fact, a small training set will not be adequate for algorithm learning; a small gallery set will make the identification too easy; and a small impostor probe set will lead to a limited FAR range. Furthermore, existing databases have limited camera views, which usually lead to a single-view probe set. To the best of our knowledge, for a specified person, no more than two cameras are related in most datasets (3D-PES has some persons seen from 3 different cameras), although the maximum number of cameras is already eight. Therefore, we collected a new database to better support the open-set person re-identification benchmark, with a sufficient training set, larger gallery and probe sets, and diverse camera views for the probe set to challenge the matching. Note that having five camera views of a person does not make the problem easier. On the contrary, this makes the problem more difficult because we put only one view in the gallery and keep all the other views independently (not associated a priori) in the probe set to make it diverse for open-set re-identification.

In the literature, several papers have addressed some similar problems of open-set person re-identification [16–19]. In [16–18], the gallery set was used as the training set, and classifiers were learned to model the known persons. Then, using the learned classifiers, a probe image can be classified as a known person

or an outlier. This is an open-set person re-identification issue, however, they did not consider to learn a general model independent of the training set, so the learned models cannot be applied to unknown galleries. Li et al. [20] proposed an open-set person verification scenario, where only binary classification was considered, namely, whether two test images belong to the same person or not. Recently, Layne et al. [19] proposed an open-world person re-identification scenario, where a flying drone was used to patrol an area, collect data, and re-identify persons. Three tasks were proposed in [19]: watch-list, intra-flight re-identification, and inter-flight re-identification. While the intra-flight and inter-flight re-identification tasks are more like a continuous single-camera or multi-camera tracking problem, the watch-list task (identify a certain watch-list person from the patrol video) is more interesting, and closely related to the open-set forensic search problem addressed in this paper. For evaluation, the conventional information-retrieval style metrics are used in [19], namely, the rank of the true matches, and the precision-recall curves. However, the rank metric is not integrated with the precision metric; therefore, the two independent metrics make the performance comparison not intuitive. In contrast, we utilize the well established open-set face recognition performance evaluation method [3–5] for the open-set person re-identification evaluation in this paper, where the rank of the true matches and the false accept rate are integrated to form a unified metric.

3   Database Collection

Our database collection was performed in a real outdoor surveillance scenario (institute campus) covered by a network of six distributed cameras, with the camera network setting illustrated in Fig. 1, where cameras have non-overlapping field of views (FOVs). The FOVs selected correspond to common pedestrian walkways, with different camera viewpoints and scene illumination conditions. Therefore, a person's appearance vary from FOV to FOV when the person walks through the monitored areas. We utilized three kinds of cameras to collect the video. For Cam1, Cam2, Cam4 and Cam5, videos were collected by the HIKVISION DS-2CD864FWD-E IP camera. For Cam3, we utilized the SAMSUNG SNB-6004P IP camera. The SAMSUNG SNB-7000P IP camera was used for Cam6. We set the image resolution to 1280×720, and the frame rate is 25fps for all the six cameras. In order to capture clear images of pedestrians with more details, we set the camera focal length to the largest and use auto iris. Nevertheless, under-exposure and occlusions may exist as we capture the videos at busy time in a cloudy day. Finally, we obtain a 125-minute video from each camera.

We further segmented two sessions of video clips from the collected videos, with one session intended for training, and the other one for test. The two sessions are about 25 minutes apart. Each session contains six video clips of about 20 minutes, with each video corresponding to one camera view. From the 12 video clips, we segmented 7413 images of 200 persons using an interactive labeling software developed by [21]. The statistics of these images are summarized in

Fig. 1. Illustration of the camera setup. The white ellipse areas indicate the FOV of a single camera, and the red trapezoidal blocks denote the positions where we locate the cameras. Six snapshots of the video frames are shown near the corresponding camera views to illustrate the captured scenes.

Table 2. The resolutions of these segmented images varying from 16×45 to 135×360.

Table 2. Number of persons and images in each camera view of each session.

	Training session		Test session	
	#persons	#images	#persons	#images
Camera 1	41	468	58	1,026
Camera 2	8	166	13	525
Camera 3	52	697	92	1,392
Camera 4	59	1,000	74	1,033
Camera 5	34	283	39	329
Camera 6	23	272	17	222
Total	81	2,886	119	4,527

One advantage of the collected database is that a person may have up to 5 camera views. This is counted in Table 3. Note that each of the 200 person has at least 2 associated camera views.

All segmented images are scaled to 128×48 pixels. Fig. 2 shows some example pairs of images from the collected database. It can be observed that there is a large variation in the observed color, and there are also lighting changes and viewpoint changes that challenge the matching of persons across cameras. Be-

Table 3. Number of persons that have multi camera views.

	Training session	Test session
2 camera views	33	70
3 camera views	43	43
4 camera views	3	6
5 camera views	2	0

sides, though all videos we recoded are high definition videos in resolution, there still exist low resolution and blur pedestrian images due to long surveillance distance.

Fig. 2. Example images from the collected database. Images in the same column represent the same person. The four rows correspond to the camera views of 1, 3, 4, and 5, respectively.

4   Benchmark Protocol

4.1   Experimental Setting

With the collected database, we design the open-set re-identification experiments as 10 random trials including training and test, and each trial involves about half of the data. The detailed procedure is described as follows.

Training Training is performed with the data of the training session. For each trial, half of the persons under each camera are randomly selected. If a person in a camera is selected, all images of that person within the camera can be used for training. Both within-camera pairs and between-camera pairs of samples can be used for learning the similarity metric. However, it is not allowed to learn a specific camera-to-camera transition model, and apply this model to the same camera pairs for testing. Though our test data has the same camera setting as the training data, we still require learning a general camera independent person re-identification model, so as to test the algorithm's generalization ability to possibly unknown camera pairs. This is also preferred in practical applications, because it is not easy to re-train the algorithm every time when dealing with new camera pairs. Note that if learning based dimension reduction method (e.g. PCA) is applied, it should be trained in each of the 10 trials for fair comparison.

Test Test is performed by applying the learned model on the test session data. For each trial, also half of the persons under each camera are randomly selected. The difference to the training data selection is that, if a person in a camera is selected, the same person in all other cameras is also selected for test. If a person is selected, all images of that person are used for test. For each trial, the learned model of the corresponding trial is applied to the randomly selected test data, and the score of each between-camera sample pair is calculated for performance evaluation.

Parameter Selection If one algorithm needs to determine the optimal hyper parameters, a cross validation procedure can be done with the training session data. Alternatively, the 10 random trials mentioned in the training procedure can also be utilized, and use the remaining half data of each trial for validation. However, any parameter tuning with the test session data is not allowed. This is to prevent hyper parameter optimization with the test data. A good algorithm should have a good generalization ability of both models and hyper parameters on unseen test data.

4.2   Performance Evaluation

In the performance evaluation, the open-set person re-identification scenario is considered. As indicated in [3], the open-set identification task is a more general scenario, with the closed-set identification being its special case. Two sub-tasks,

detection and identification, are involved in the open-set identification process. In the detection sub-task, the system decides whether the person of the probe image presents in the gallery or not. In the identification sub-task, the system reports the identity of the accepted probe. Therefore, the task of open-set identification is to determine the identity of the probe or to reject the probe.

The performance evaluation of the open-set person re-identification task involves three sets of images. The first set is the gallery set G, which contains person images collected by the system (e.g. pedestrians detected and segmented from a video). The other two are probe sets P_G and P_N. While P_G consists of persons in the gallery set G but with different images, P_N includes persons that are not present in G. Two performance measures, the detection and identification rate (DIR), and the false accept rate (FAR), are calculated for evaluation [3–5]. Let $id(g,p)$ be an indicator whether g and p belong to the same identity, that is,

$$id(g,p) = \begin{cases} 1, & g \text{ and } p \text{ belong to the same identity,} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Let $s(\cdot,\cdot)$ be the similarity score function. Let

$$g^* = arg \max_{g \in G, id(g,p)=1} s(g,p), \tag{2}$$

that is, g^* is the gallery image that has the same identity as the probe image p and reaches the maximum score among all gallery images of the same identity. Furthermore, let $rank(p)$ denote the rank order of $s(g^*,p)$ among matching scores between p and all gallery images. That is, $rank(p) = k$ means that $s(g^*,p)$ is the kth largest similarity score. Then, the DIR and FAR measures are formulated as

$$DIR(\tau,k) = \frac{|\{p|p \in P_G, \ rank(p) \leq k, \ s(g^*,p) \geq \tau\}|}{|P_G|}, \tag{3}$$

$$FAR(\tau) = \frac{|\{p|p \in P_N, \ \text{and} \ \max_{g \in G} s(g,p) \geq \tau\}|}{|P_N|}, \tag{4}$$

where τ is the decision threshold, and $|A|$ calculates the number of elements in the set A.

Given a rank level k, by changing the decision threshold τ, a Receiver Operating Characteristic (ROC) curve can be drawn by plotting DIR versus FAR. Given an FAR level, the operating threshold τ can be determined by Eq. (4), and a CMC curve of DIR versus the rank k can also be plotted. Note that when FAR=100%, the corresponding DIR is the traditional identification rate of the closed-set identification task, with the gallery set G and the probe set P_G.

In the evaluation of the open-set person re-identification, for each trial, we construct the three sets G, P_G and P_N using all the randomly selected test images. First, for a camera c, a gallery set G_c is constructed using all the selected test images from that camera. Second, images from other cameras with persons belonging to G_c are used to construct a probe set P_{G_c}, and all other images

form the other probe set P_{N_c}. The three sets are used to evaluate the open-set person re-identification performance measures DIR_c and FAR_c according to Eqs. (3) and (4). Then, each camera is used to construct a gallery set in tune, and all performance measures DIR_c and FAR_c are averaged, respectively, to get the mean performance of the trial. Finally, performance measures of all the 10 random trials are averaged.

Furthermore, we also compute the standard deviation over the 10 trials of the performance measures, and we adopt the fused performance measure $\mu - \sigma$ as proposed in [5], which considers both the mean and standard deviation statistics for performance evaluation. The purpose of the fused measure is to enforce comparison of the standard deviation. This metric requires algorithms to be stable in all random trials.

The above evaluation procedure is illustrated in Fig. 3.
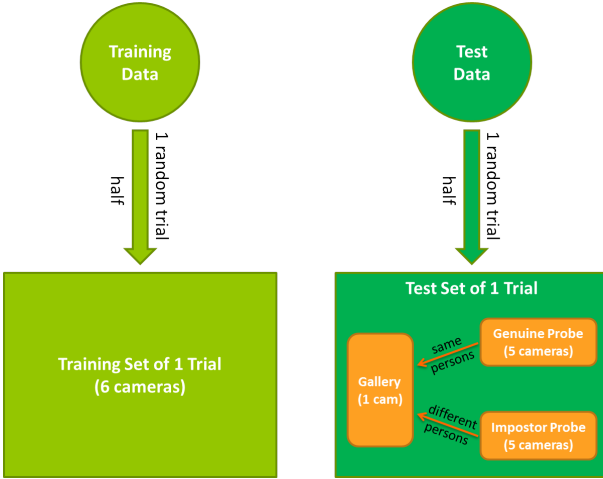
Fig. 3. Illustration of the OPeRID performance evaluation. For each trial, half of the training data is randomly sampled to learn a general distance or similarity metric, and half of the test data is also randomly sampled for performance evaluation. In addition, the randomly sampled test data is divided into the gallery set G with a certain camera, the genuine probe set P_G containing the same persons as in G but with the other five cameras, and the impostor probe set P_N containing different persons from G who appear in the other five cameras. This test set partition is repeated for six times where each time a gallery set is constructed with one of the six cameras in tune, and the performance is averaged. The above procedure is repeated for 10 trials, and the mean and standard deviation of the performance is calculated.

5   Baseline Methods

5.1   Feature Representation

For feature representation, we utilize a sliding window based histogram representation. Three scales of subwindows, 16×16, 32×32, and 48×48, have been used. The sliding step is set to be 8 pixels. In a 128×48 image, this generates 125 subwindows in total.

In each subwindow, we extract both color and texture features. For color features, the HSV color space is adopted. The H channel is quantized as 8 bins, while the S and V channels are quantized as 3 bins. Then, a joint histogram of $8 \times 8 \times 3$ bins is extracted in each subwindow. For the texture feature, we adopt the Multi-scale Block based Local Binary Patterns (MB-LBP) [22]. Three scales of MB-LBP, 3×3, 7×7, and 11×11, have been used. A rotation invariant encoding [23] is applied, resulting a 30-bin histogram in a subwindow for the three scales of MB-LBP. The color and texture histograms of all subwindows are concatenated to build the final descriptor. To suppress the influence of very large feature values, we truncated each histogram bin values to be no larger than 255, and then made a square-root transform. We make all the extracted features from our collected database publicly available in the project website, so that future classifier or metric learning algorithms can be applied on the same feature set for evaluation.

The dimensions of our feature descriptor is 12,750. We apply the Principle Component Analysis (PCA) approach in each trial to reduce the dimensions to 100 for further metric learning.

5.2   Metric Learning

We used a software package provided by [24] to evaluate several metric learning algorithms, including Identity (i.e. Euclidean distance), MAHAL (i.e. Mahalanobis distance learned from positive sample pairs), LMNN [25], ITML [26] and KISSME [24]. The LMNN algorithm [25] aims at learning a Mahalanobis distance metric for improving the k-nearest neighbor (kNN) classification, which can be solved by semidefinite programming. The ITML approach [26] considers minimizing the differential relative entropy between two multivariate Gaussians for learning the Mahalanobis distance function. The KISSME algorithm [24] applies the log likelihood ratio test of two Gaussian distributions for metric learning, and so a simplified solution can be derived. Besides, we also utilized the LADF algorithm (Locally-Adaptive Decision Functions) [27] for evaluation. This algorithm can be viewed as a joint model of a distance metric and a locally adapted thresholding rule. Furthermore, we implemented a linear regression based discriminant analysis method introduced in [28]. This method learns a discriminant subspace via linear regression of class labels. Different from [28], we applied the ridge regression method instead of the least square regression for robust learning. We call the resulting algorithm Ridge Regression based Discriminant Analysis (RRDA). We further used the Cosine similarity metric on the derived subspace to measure the similarity between two pedestrian images.

6   Baseline Performance

In this section, we summarize the baseline performance of the open-set person re-identification problem, evaluated on our collected multi-camera database. We follow the procedure described in the benchmark protocol, resulting in the ROC curves with rank=1 and rank=10 shown in Figs. 4 and 5. Recall that FAR=100% corresponding to the traditional closed-set person re-identification problem. From the figures it can be seen that the closed-set person re-identification performances of the evaluated algorithms are promising, compared to the performance observed from other databases. However, it is clear that the open-set person re-identification performance significantly drops with the decreasing value of FAR. This means that the current algorithm is still far from satisfactory in the real application aspect of view. Nevertheless, the simple RRDA algorithm is shown to be quite superior for open-set person re-identification, compared to other evaluated algorithms.
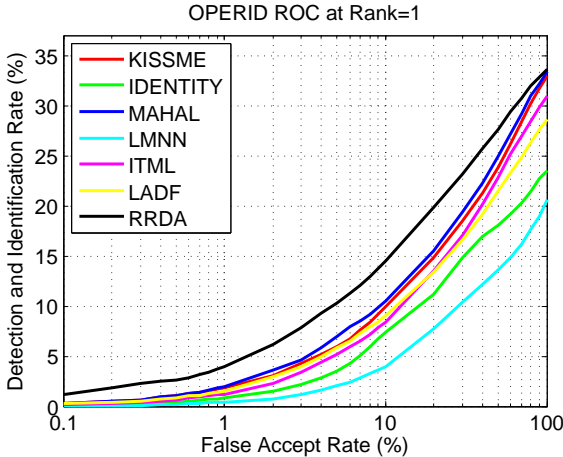
Fig. 4. ROC curves with rank=1. Performances are measured in $\mu - \sigma$ of 10 trials.

We also calculated the performance at fixed FARs, and generated CMC curves under FAR=1% and FAR 10%, as shown in Fig. 6 and Fig. 7. From these curves it is observed that all algorithms perform very poor at the two FAR points. What is worse, it can also be observed that increasing the number of retrievals after rank 10 helps very little in improving the performance. This is because genuine matching scores at lower ranks are hardly be larger than the decision threshold.

To be specific, we summarize some results in Table 4. The results clearly indicate a demand for improving the open-set person re-identification performance, since all algorithms did not exceed 17%. Nevertheless, it can be observed that RRDA is the best performer among the evaluated algorithms. In summary,
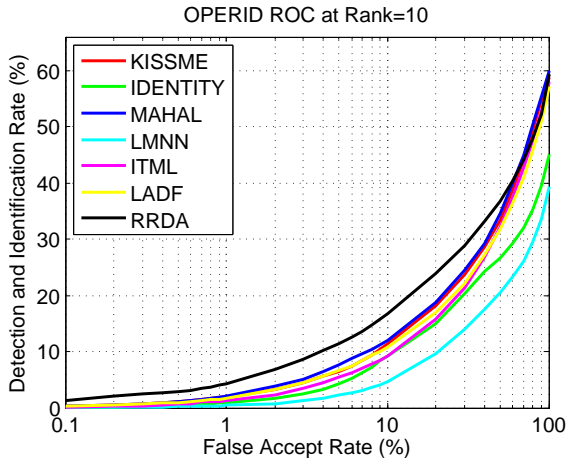
Fig. 5. ROC curves with rank=10. Performances are measured in $\mu - \sigma$ of 10 trials.

the open-set person re-identification problem is still far from solved, and deserve further research and development.

Table 4. Detection and identification rates (%) of metric learning algorithms under several operating points. Performances are measured in $\mu - \sigma$ of 10 trials.

	FAR=1%		FAR=10%	
	Rank=1	Rank=10	Rank=1	Rank=10
IDENTITY	0.84	0.91	7.36	9.21
MAHAL [24]	1.89	1.99	10.50	11.97
KISSME [24]	1.82	1.92	9.99	11.46
LMNN [25]	0.41	0.41	3.97	4.58
ITML [26]	1.18	1.21	8.39	9.27
LADF [27]	1.53	1.74	9.11	10.82
RRDA	**3.99**	**4.35**	**14.51**	**16.72**

7   Conclusions

In this paper, we have considered the open-set person re-identification problem, which includes two sub-tasks, detection and identification. We have contributed a database collected from a video surveillance setting of 6 cameras, with 200 persons and 7,413 images segmented. Based on this database, we have developed a benchmark protocol for evaluating the performance under the open-set person re-identification scenario. Several popular metric learning algorithms for person re-identification have been evaluated as baselines. From the baseline performance, we conclude that the open-set person re-identification problem is still
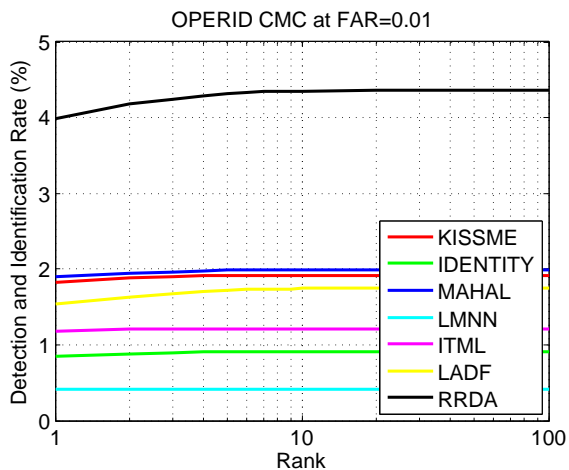
Fig. 6. CMC curves with FAR=1%. Performances are measured in $\mu - \sigma$ of 10 trials.
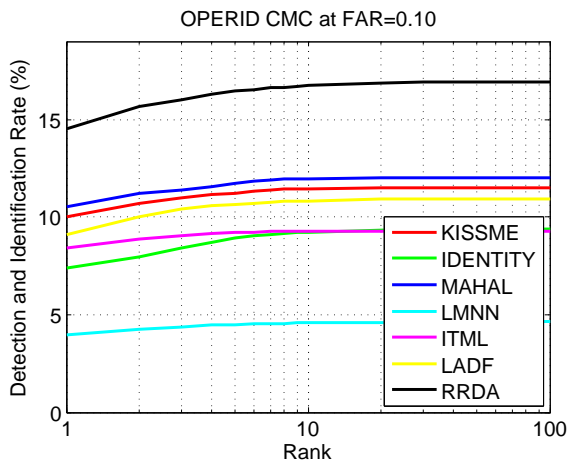
Fig. 7. CMC curves with FAR=10%. Performances are measured in $\mu - \sigma$ of 10 trials.

largely unresolved, thus further attention and effort is needed. We have released the collected database OPeRID v1.0, as well as the extracted features and the benchmark tool to advance research along this direction.

References

1. Vezzani, R., Baltieri, D., Cucchiara, R.: People reidentification in surveillance and forensics: A survey. ACM Computing Surveys (CSUR) **46**(2) (2013) 29
2. Gong, S., Cristani, M., Yan, S., Loy, C.C.: Person Re-Identification. Springer (2014)
3. Phillips, P.J., Grother, P., Micheals, R.: Evaluation methods in face recognition. In Li, S.Z., Jain, A.K., eds.: Handbook of Face Recognition. 2nd edn. Springer (Sep. 2011) 551–574
4. Liao, S., Jain, A.K., Li, S.Z.: Partial face recognition: Alignment-free approach. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(5) (2013) 1193–1205
5. Liao, S., Lei, Z., Yi, D., Li, S.Z.: A benchmark study of large-scale unconstrained face recognition. In: IAPR/IEEE International Joint Conference on Biometrics, Clearwater, Florida, USA (Sep. 29 - Oct. 2 2014)
6. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: IEEE International workshop on performance e-valuation of tracking and surveillance. (2007)
7. Ess, A., Leibe, B., Van Gool, L.: Depth and appearance for mobile scene analysis. In: Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, IEEE (2007) 1–8
8. Schwartz, W.R., Davis, L.S.: Learning discriminative appearance-based models using partial least squares. In: Computer Graphics and Image Processing (SIB-GRAPI), 2009 XXII Brazilian Symposium on, IEEE (2009) 322–329
9. Nilski, A.: Evaluating multiple camera tracking systems-the i-lids 5 th scenari-o. In: Security Technology, 2008. ICCST 2008. 42nd Annual IEEE International Carnahan Conference on, IEEE (2008) 277–279
10. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE (2011) 649–656
11. Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: BMVC. Volume 2. (2011) 6
12. Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person re-identification by de-scriptive and discriminative classification. In: Proc. Scandinavian Conf. on Image Analysis. (2011)
13. Loy, C.C., Xiang, T., Gong, S.: Multi-camera activity correlation analysis. In: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE (2009) 1988–1995
14. Baltieri, D., Vezzani, R., Cucchiara, R.: 3dpes: 3d people dataset for surveillance and forensics. In: Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding, ACM (2011) 59–64
15. Li, W., Wang, X.: Locally aligned feature transforms across views. In: IEEE Conference on Computer Vision and Pattern Recognition. (2013)
16. Bauml, M., Stiefelhagen, R.: Evaluation of local features for person re-identification in image sequences. In: IEEE International Conference on Advanced Video and Signal-Based Surveillance. (Aug. 2011)

17. Bedagkar-Gala, A., Shah, S.: Multiple person re-identification using part based spatio-temporal color appearance model. In: IEEE International Conference on Computer Vision Workshops. (Nov. 2011)
18. Eisenbach, M., Kolarow, A., Schenk, K., Debes, K., Gross, H.: View invariant appearance-based person reidentification using fast online feature selection and score level fusion. In: IEEE International Conference on Advanced Video and Signal-Based Surveillance. (Sep. 2012)
19. Layne, R., Hospedales, T.M., Gong, S.: Investigating open-world person re-identification using a drone. In: ECCV Workshop on Visual Surveillance and Re-identification, Switzerland (2014)
20. Li, A., Liu, L., Wang, K., Liu, S., Yan, S.: Clothing attributes assisted person re-identification. Circuits and Systems for Video Technology, IEEE Transactions on (2014)
21. Vondrick, C., Patterson, D., Ramanan, D.: Efficiently scaling up crowdsourced video annotation. International Journal of Computer Vision **101** (January 2013) 184–204
22. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z.: "Learning multi-scale block local binary patterns for face recognition". In: Proceedings of the 2nd IAPR/IEEE International Conference on Biometrics. Volume LNCS-4642., Seoul, Korea (2007) 828–837
23. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence **24**(7) (2002) 971–987
24. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: IEEE Conference on Computer Vision and Pattern Recognition. (2012)
25. Weinberger, K., Blitzer, J., Saul, L.: Distance metric learning for large margin nearest neighbor classification. Advances in neural information processing systems (2006)
26. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: Proceedings of the 24th international conference on Machine learning, ACM (2007) 209–216
27. Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R.: Learning locally-adaptive decision functions for person verification. In: IEEE Conference on Computer Vision and Pattern Recognition. (2013)
28. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer (2009)