Investigating Open-World Person Re-identification Using a Drone

Ryan Layne, Timothy M. Hospedales, and Shaogang Gong

Vision Group, School of EECS, Queen Mary University of London, UK

Abstract. Person re-identification is now one of the most topical and intensively studied problems in computer vision due to its challenging nature and its critical role in underpinning many multi-camera surveillance tasks. A fundamental assumption in almost all existing re-identification research is that cameras are in fixed emplacements, allowing the explicit modelling of camera and inter-camera properties in order to improve re-identification. In this paper, we present an introductory study pushing re-identification in a different direction: re-identification on a mobile platform, such as a drone. We formalise some variants of the standard formulation for re-identification that are more relevant for mobile re-identification. We introduce the first dataset for mobile re-identification, and we use this to elucidate the unique challenges of mobile re-identification. Finally, we re-evaluate some conventional wisdom about re-id models in the light of these challenges and suggest future avenues for research in this area.

1 Introduction

Person re-identification has been extensively and aggressively studied in recent years by the computer vision community due to its challenging nature and critical role in underpinning many security and business-intelligence tasks in multi-camera surveillance [9]. This has resulted in continued improvements in performance on increasingly challenging benchmark datasets. In essence re-identification is about successfully retrieving people by *identity*, enabling security operators or higher-level software components to locate individuals. Nevertheless, it is conventionally formulated as a one-to-one set-matching problem between two fixed cameras, for which an effective model can be learned. In this paper we present an introductory study that relaxes this core assumption and investigates how re-identification generalises to mobile surveillance platforms as realised by quadrocopter drones [5].

Despite the successes of static CCTV cameras, we argue that considering alternative surveillance equipment not only opens up exciting new research areas, but also new ways of thinking about re-identification and particularly, how re-identification fits into real-world applications and links with other research fields. New technology such as remotely-operated vehicles and wearable visual sensing equipment is becoming increasingly accessible in terms of cost and availability to the general public. In many cases, quickly deployable mobile visual systems rival

currently predominant static CCTV cameras in terms or resolution and frame-rate. More critically, they intrinsically have a qualitative flexibility advantage – in terms of being mobile – and are thus able to dynamically adapt their viewing position and direction without being constrained by the emplaced locations of a CCTV camera. We term any piece of equipment that can perform video surveillance in a portable sense, a *mobile re-identification platform* or, MRP.

While generalising re-identification to MRPs provides many new capabilities and research avenues, it introduces some significant differences and new challenges compared to the standard formulation of the re-identification problem. These broadly relate to the interrelated issues of (1) view ambiguity, (2) view variability and (3) open-world re-id.

Within-view ambiguity: The first major contrast between MRP and standard fixed camera re-id relates to the number of views. That is, the standard setting is typically defined across a pair of camera views, and within-camera tracking is typically assumed to fully disambiguate detections within-view. In contrast for MRPs 'within camera' re-id is itself non-trivial because the camera's positional and orientational mobility means that even stationary people frequently enter and exit the view area due solely to self-motion of the platform. This further generalises the so called 'MvsAll' scenario described in [14] to 'AllvsAll'.

View Variability: The second major contrast is the continually varying view-stream of a MRP compared to the conventional fixed position CCTV camera. This is significant because most of the recent performance gains in the state of the art re-id methods have come from supervised learning of *view* or *view-pair specific* models [10]. In the MRP case the *continually* varying view parameters – including range, lighting, self induced motion blur and detection alignment – precludes learning such models (see Figure 1).

Open-world: Most existing re-id studies make the simplifying assumption of closed-world conditions. That is that there is a one-to-one set match, where everyone in the first camera re-appears in the second camera. No one disappears, and no extra people appear. Although convenient for modelling and benchmarking purposes, this is clearly an extremely strong assumption in practice. In the case of MRP with within-camera re-id ambiguity, and the mobile nature of the platform, closed-world is clearly an inappropriate assumption – meaning that re-id with MRP is significantly more ambiguous than the conventional setting.

Despite the challenges identified above, MRPs provide a compelling new ground to break for re-identification science both in terms of broadening the application area as well as providing the opportunity to reconsider several implicit but strong assumptions made in most existing re-id research. In this work, we make four main contributions: (i) We present a case for the pursuit and development of a new research area using mobile re-identification platforms (MRPs); (ii) We formalise three novel MRP-related variants on the classic re-identification scenario; as well as associated evaluation metrics for each; (iii) We collect the first public dataset for MRP re-id and establish benchmarks for each of the iden-

Fig. 1. Illustrating key differences in person detection quality when automatically detected from mobile re-identification platform video (MRP, left), compared to detections in a standard re-identification dataset, VIPeR (right). Notably, the VIPeR images (i) are in perfect register, (ii) feature standard walking poses from a limited number of relative angles. Contrastingly, the MRP images are unregistered, feature more varied pose and also occasionally heavy motion-blur because of the relative motion of the MRP to the target person during transit.

tified tasks; (iv) We elucidate the unique challenges posed by MRP re-id and discuss their implications for general re-id research going forward.

2 Related Work

Re-identification: There is now an extensive body of research on conventional re-identification, broadly split into contributing effective feature representations [7, 33], discriminative matching models [3, 1], or both [18]. The other major design axis typically considered is 'single-shot' [3, 1, 33] (exactly one image per person) versus 'multi-shot' [7, 15, 26] (exploiting multiple images per person where available to improve results). For a broad background of research to this paper we suggest [10] and [31]. Going beyond conventional re-identification, we next discuss a few recently identified research areas that are relevant to our MRP context.

Open-world Re-Id: At its most general, open world re-identification [9] addresses relaxing several assumptions: one-to-one set-match (that is, that every person in the probe set appears in the gallery set and vice-versa) [13]; the assumption of matching between only two cameras [13]; the assumption of a known number of people; or the assumption that multi-shot grouping is known a-priori [14]. A few studies have begun to work toward this including [13, 14]. However, these have generally considered only a couple of these relaxations at once. In contrast, the MRP re-id scenario is intrinsically open-world: self movement in a potentially open-space means one-to-one match situations are unlikely, self-motion means that tracking cannot provide multi-shot grouping, and clearly the person count of an arbitrarily surveilled space is not known in advance.

Generalised-view Re-Id: The conventional approach to maximising re-id performance is learning a discriminative model to maximise re-id rate for a specific pair of fixed camera views [3, 1]. A few studies have started to consider how re-identification models generalise across views [19] and generally found that they don't – achieving good re-id rate requires view specific discriminative training. This reflects analogous conclusions drawn more broadly in computer

vision recognition [30]. As a result, studies have begun to develop transfer strategies that allow models learned from 'source' view pair(s) to be adapted to better apply in a new 'target' view [4, 19, 22] which may have different position, lighting, etc. These studies have generally considered combining [4] or adapting [19, 22] source model(s) to construct the model for a new domain – with the general aim of reducing or eliminating the need for collecting annotated training data for every pair of cameras. The important contrast with our MRP context is that domains/camera pairs as described above are *discrete*. In contrast, the video feed from a MRP is a *continuously varying* domain. This means that for previous approaches to view generalisation it is still assumed that enough data to model a specific view or view pair can be collected and a discriminative model learned. This is no longer feasible for MRP, since the constantly varying view means that collecting (let alone annotating) extensive view-specific data is impossible, and the conventional strategy of learning a discriminative model is called into question.

Drones: A full discussion of background research in drone technology is out of the scope of this paper, but see [5] for an introduction and background to drones and their capabilities. The central issue for drones to become more useful for surveillance tasks is for them to become increasingly autonomous, and a significant component of this is learning to maintain consistent person identity estimates over time, which we address here.

3 Re-identification Problem Variants and Metrics

Conventional re-identification is used as a forensic search tool, or as a module by higher-level software – such as inter-camera tracking [25]. For ease of model formulation (e.g., metric learning, SVM ranking), evaluation and establishing benchmarks, most studies formalise re-id as a closed-world set match between two specific cameras. As a result the typical evaluation metric is Rank 1 accuracy (the % of perfect gallery matches for each probe image), or the CMC curve (the % of correct matches within the top N ranked matches, for varying N) [32]. In this section we describe three distinct variants of the re-identification problem that naturally arise with MRPs – each based on intuitive application scenarios for a MRP. Table 1 summarises the problem variants proposed and compares them with classical approaches to re-id.

3.1 Watchlist Verification

In the *watchlist* task, the MRP is patrolling an area and the goal is to detect if any person encountered is somebody on a pre-defined watch-list. For the moment we make no assumption on whether the MRP is manually controlled, has a pre-programmed travel path or autonomously wanders. However, we assume that the scenario is *passive sensing* – the MRP is not going to to take action based on any detected matches. The watchlist itself could come from a variety of sources: a pre-defined mug-shot gallery; a transmitted detection from another MRP or

CCTV camera; or a previous detection saved by the current MRP on a previous flight or earlier in this patrol. For example the MRP may be trying to track down a specific person previously identified performing a suspicious action of interest.

In this case, the 'probe' is a single person from the watch list, and the 'gallery' is all people observed in a patrol. In contrast to conventional re-id, this is a more open world problem in that: (i) the probe person may not appear anywhere in the patrol video (no match is an option), (ii) (most) people in the patrol video are not on the watchlist (many background distractors), and (iii) the total number of detected instances of the true match if present in the gallery/patrol video is unknown (not one-to-one). In Table 1 this is illustrated under match by [N] and [M] reflecting multiple potential *ungrouped* matches and distractors respectively.

Given these considerations, the right evaluation metrics for this problem are information-retrieval style metrics, thus we use a suite of them: (i) the rank of the true matches, and (ii) precision-recall curves and associated summaries – average-rank and average-precision.

3.2 Within-Flight Re-Identification

In the *within-flight* re-identification task, the MRP's goal is to maintain consistent identity of person detections recorded throughout the flight. Due to both platform and target motion, a particular target may enter the view once, or enter and exit the view multiple times throughout the flight. In this case there is only one "camera view" as compared to conventional re-id setting of two fixed cameras. However, it means that: (i) the platform motion can create potentially more view-variation over time than occurs between two fixed CCTV cameras, so "within-view" re-identification can become even harder than conventional re-id; (ii) as before, there is a general open-world identity inference problem.

The general identity inference problem here means that there is no-longer a notion of probe and gallery. Instead there is a list of N detections, which each need to be assigned one of $K \leq N$ unique identities. However K (the number of unique people in the scene) is itself unknown. In Table 1 this is illustrated under match by [N] – the single set of detections with unknown grouping – and an unknown person count.

Evaluating this open world identity assignment is non-trivial compared to closed world. To fully evaluate the performance, we use statistical analysis on all pairs of detections to measure pairwise Precision and Recall. Specifically given all true $\mathcal{L}_{gt}$ and estimated $\mathcal{L}_{est}$ labels of the N detections. A 'true' pair i, j has the same label, and a 'false' pair have different labels. Thus true-positive, true-negative, false positive and false-negative rates can be computed as in Eq. (2); which can in turn be summarised in terms of Precision, Recall, Specificity, and Accuracy as in Eq. (1).

Setting	Cameras	Match	Person Count	View-specific	Multi-shot	Evaluation
Singleshot [7, 33, 3, 1]	2	$N:N$	Known	Yes	No	Rank 1, CMC
Multishot [7, 15]	2	$N:N$	Known	Yes	Grouped	Rank 1, CMC
Karaman [14]	2	$N:[N]$	Known	Yes	Group : No group	Accuracy
John [13]	2	$N+M_1:N+M_2$	Known	Yes	No	Rank 1
Watchlist	1	$1:[N]+[M]$	N/A	No	No group	Rank, Prec+Recall
Within	1	$[N]$	Unknown	No	No group	F-measure
Across	2	$[N]+[M_1]:[N]+[M_2]$	Unknown	No	No group	F-measure

Table 1. Contrasting re-identification problem variants. Match: $N:N$ reflects closed world one-to-one mapping among N people in view 1 : view 2. $[N]$ indicates unknown within-camera grouping. M represents the unknown fraction of the people to be matched who are distractors in that they do not occur in the other view or the watchlist.

$$\begin{aligned} Prec &= TP/(TP+FP) \\ Rec &= TP/(TP+FN) \\ Spec &= TN/(FP+TN) \\ Acc &= (TP+TN)/N \end{aligned} \quad (1)$$

$$\begin{aligned} TP &= \sum_{ij} (\mathcal{L}_{gt}(i) = L_{gt}(j)) \wedge (\mathcal{L}_{est}(i) = \mathcal{L}_{est}(j)) \\ TN &= \sum_{ij} (\mathcal{L}_{gt}(i) \neq L_{gt}(j)) \wedge (\mathcal{L}_{est}(i) \neq \mathcal{L}_{est}(j)) \\ FP &= \sum_{ij} (\mathcal{L}_{gt}(i) \neq L_{gt}(j)) \wedge (\mathcal{L}_{est}(i) = \mathcal{L}_{est}(j)) \\ FN &= \sum_{ij} (\mathcal{L}_{gt}(i) = L_{gt}(j)) \wedge (\mathcal{L}_{est}(i) \neq \mathcal{L}_{est}(j)) \end{aligned} \quad (2)$$

3.3 Across-flight re-identification

The *across-flight* problem is somewhat more related to the classic problem of between-camera re-id. In this case identities should be matched across two separate MRP flights. This may be from either the same platform making two patrols, or two distinct and communicating platforms trying to coordinate identities. It is a fully open-world problem, given that within-flight/view tracking cannot be assumed for MRPs (ungrouped detections in Table 1), and that only an unknown subset of the total people in each view may be shared (in Table 1, N shared + M distractor people in each view). However, compared to within-flight re-identification, it may be somewhat harder because the environments across space and/or time may be even more different than the view change caused by platform motion in the previous case. Again, statistical analysis is the appropriate evaluation technique.

4 Methodology and Experimental Setting

4.1 Data acquisition

Drone Setup We use a standard remote-operated quadrocopter to realise our MRP for data acquisition. During data collection, a human operator controlled

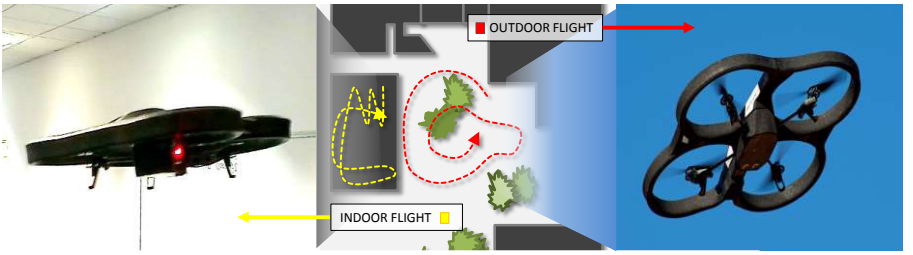

Fig. 2. Flight path detail (center) and images of the drone used in our experiments indoors (left) and outdoors (right).

the drone via laptop using the Robot Operating System (ROS[1]) to ensure responsive handling with the control loop and sensor data capture operating at ≈ 200Hz whilst video from the quadrocopter was sampled at $\approx 1-5$Hz. For this particular commodity platform, flight time was limited by battery capacity to ≈ 10 minutes per flight at 640x360 pixels.

During flight, a heads-up-display (HUD) is overlaid on top of the video feed displaying standard sensor information (such as yaw, pitch, acceleration, battery and altitude), as well as real-time person detections and person detection confidence scores. This in some sense serves to provide the operator with the visual cues necessary to weakly simulate an active-sensing, fully autonomous (i.e. closed-loop) drone. If the drone is orientated poorly towards a person or the person is partially occluded then a poor detection will result and the operator can adjust the relative orientation and position of the drone based on this visualisation until a strong detection can be obtained. Some examples of the HUD can be seen in Figure 3.

Person detection Given the $1-5$Hz video feed, the next task is to obtain person detections. To maximise the reliability of this step, we first apply a corrective transform on each frame to correct for the 'roll' of the drone (using data recorded from the MRP's onboard accelerometer sensor), since the detection models assume people to be upright. In order to detect people fast enough for real-time visualisation so as to assist the MRP's operator, we employ [6]'s toolkit which provides excellent computational efficiency and detection quality. At extraction time, we resample detections to [128x48] pixels[2]. We threshold detections and discard any with a confidence of below 20% since the environments from which we will be detecting are extremely varied with respect to lighting and pose and we wish to limit the number of potential false-positive detections whilst retaining most true detections. For our visual features we employ the commonly used

[1] http://www.ros.org/

[2] However, note that the original resolution and therefore resample quality will vary dramatically over time within a flight, see Figure 1.

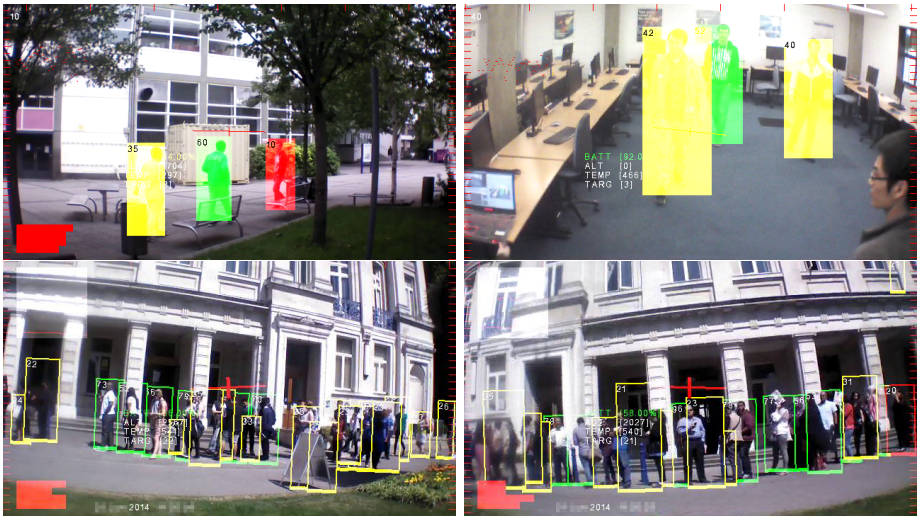

Fig. 3. Screen captures from our mobile re-identification platform's data capture sessions; illustrating real-time person detections colour-coded by detection confidence. The top-left and top-right images illustrate typical operator views from the outdoor and indoor flights from Dataset 1; The bottom row illustrates Dataset 2.

ensemble of local features [11] (ELF), which encodes both color and texture in 6 horizontal strips [24] for final features of 2784 dimensions.

Datasets Using the procedure described above, we collected two multi-flight datasets. **Dataset 1:** The first dataset contains three flights worth of data, across an outdoor and indoor environment. These consisted of 436, 652, 848 video frames, from which we obtained 233, 471, and 797 person detections from 6, 7, and 10 distinct people (after thresholding). All person detections in this dataset are exhaustively annotated. **Dataset 2:** The second, significantly larger, dataset contains six flights of data in three different unconstrained and heavily crowded outdoor environments. Across each flight there are between 10,000 and 30,000 frames of video data and an average of 8,654 person detections from an unknown number of distinct people. Of this data, we selected a single flight and exhaustively annotated 28 unique identities within the 4096 detections available within a 2:06 window[3].

4.2 Classifier training, Representation and Datasets

Training Strong Models One of the central questions we wanted to answer is to what extent the state of the art discriminative models for standard benchmark datasets are effective for MRP based re-identification. This question is crucial

[3] All datasets and annotations will be realised on the web at http://qml.io/rlayne

because conditions in MRP-sourced video data continuously change during a flight thus there are many more combinations of pose and viewing angle than in the fixed view case assumed by most state of the art models – i.e. a fixed view with enough (annotated) data is sufficient to learn a model. It is therefore critical to discover if and how much performance discriminative models lose on dynamically changing data.

We investigate this by training a selection of strong discriminative models including one of the most popular: RankSVM [24]; and two recent state of the art approaches BR-SVM [2] and KISS [17]. We train these models on a variety of large benchmark datasets including VIPER [27] (632 distinct persons in [128x48] crops), PRID [12] (200 distinct persons), GRID [21] (250 persons) and CUHK [20] (971 persons). We resample all detections to match VIPeR's dimensions. For the computationally intensive discriminative methods, we reduce the dimension with PCA to $d = 200$ for BR-SVM and $d = 34$ for KISS as specified in [17].

Domain Shift Since we assume a stationary view and the absence of live-annotation of video-feed data (as proxies for normal discriminative training on a single-view), the only way to apply trained matching models for MRPs is to train them on benchmark datasets before testing them on the MRP video feed. This potentially opens up the issue of *domain shift* [8, 23, 19] between the training and testing data. For example, due to additional chance of motion blur, mis-registered images and more variance in pose from the MRP detections (Figure 1), which are absent in VIPER.

As a preliminarily investigation into how to overcome this issue, we consider unsupervised domain-adaptation in order to better align the target MRP data X_t and source VIPeR training data X_s. That is, warp $p(X_t)$ so that it is more aligned with the source training data $p_{adapt}(X_t) \approx p(X_s)$, with the intuition that this should allow classifiers trained on X_s to generalise better to X_t [23]. In particular, we align the projected subspaces of the two datasets, using the geodesic flow kernel domain adaptation (DA) method [8] using $d_{DA} = 13$ dimensions..

4.3 Re-identification and baselines for comparison

For **Task 1: Watchlist**, we simulate this experiment by taking each person detection in turn as the watch-list, and matching it against every other detection from the flight to produce a ranked list. The ranked list of results is then evaluated for relevance with information retrieval metrics (Sec 3.1). Whether first, average or last rank; or average precision is the most relevant metric will depend on the end-user application and cost function. We evaluate this task with both Datasets 1 and 2. For **Task 2: Intra-flight re-identification** and **Task 3: Inter-flight re-identification** (Sec 3.2-3.3), the experiment is performed by matching every detection against every other detection. The resulting detection-affinity matrix is thresholded[4] and analysed for connected components [29]. Each connected component defines an estimated person. The estimated $\mathcal{L}_{set}$ and true

[4] The threshold is chosen to optimise F-measure for each model.

$\mathcal{L}_{gt}$ identities are compared using statistical analysis as explained in Section 3. We evaluate these tasks with Dataset 1. As algorithms to produce the matching scores for each experiment, we compare the following models:

NN-[DA] Nearest-neighbor (NN) matching based on the detection descriptor.
BR-SVM-[DA] Binary-relation SVM with RBF concatenation kernel [2].
RankSVM-[DA] SVM with difference feature and linear kernel [24].
KISS-[DA] State of the art discriminative Mahalanobis metric learning [17].

In each case we compare the model with and without domain adaptation (-DA suffix). As explained earlier, we do not have annotated view-specific training data. Thus, we train the latter three discriminative models on the full VIPER dataset of 632 pairs and test them on the MRP video detections. These models obtain good results when applied within-domain on VIPER [2, 24, 17], however our experiment will test their ability to generalise this knowledge to a continuously varying view.

5 Experiments

5.1 Watchlist and Re-identification evaluations

We first present the results for the three main tasks before drawing conclusions from them.
Task 1: Watchlist The results of watchlist verification are presented in Table 2(a) for Dataset 1, and Table 2(b) for Dataset 2. This task reflects how highly true matches to each particular watchlist person are ranked relative to all the other person detections in the dataset, on average. Clearly all methods perform better than random: average rank, for example, has a chance level of half the number of detections across all flights which is $500/2 = 250$ for Dataset 1 and $4046/2 = 2023$ for Dataset 2. The best methods obtain a first rank result of around 2. Surprisingly, this is the case both in the smaller Dataset 1 and the larger Dataset 2.
Task 2: Intra-flight re-identification Intra-flight re-id results for Dataset 1 are presented in Table 3(a). This task attempts un-constrained detection association across all detections within a flight.
Task 3: Inter-flight re-identification Intra-flight re-id results for Dataset 1 are presented in Table 3(b). This task attempts un-constrained detection association across all detections from a pair of flights.

5.2 Observations and Analysis

Based on the results described in the previous section and Tables 2-3, we make the following observations and conclusions.
(1) NN is best overall – Surprisingly, outperforming all discriminative meth-

Dataset 1	NN	NN-DA	KISS	KISS-DA	BRSVM [1]	BRSVM [1] DA	RankSVM [24]	RankSVM [24] DA
First rank ↓	2.08	4.69	4.15	5.37	12.32	15.87	9.76	17.53
Last rank ↓	167.93	162.89	156.70	150.82	166.35	160.78	177.32	170.37
Average rank ↓	56.30	56.47	54.45	57.39	65.65	68.77	76.81	81.51
Average Prec ↑	0.46	0.46	0.43	0.41	0.34	0.35	0.24	0.24
Dataset 2	NN	NN-DA	KISS	KISS-DA	BRSVM [1]	BRSVM [1] DA	RankSVM [24]	RankSVM [24] DA
First rank ↓	1.91	2.47	18.02	9.89	265.59	18.87	280.57	424.64
Last rank ↓	1864.34	2001.95	2152.18	2032.83	2841.16	2238.77	2673.06	3357.40
Average rank ↓	507.30	528.85	619.78	635.10	1256.77	753.53	1213.27	1848.23
Average Prec ↑	0.36	0.34	0.19	0.25	0.04	0.14	0.04	0.02

Table 2. Watchlist verification results for each model. Top: Dataset 1, results are averages over all persons and all flights, average 500.3 total detections. Bottom: Dataset 2, results are for single annotated flight, 4046 total detections. For the rank metrics lower is better (↓) and for the average precision metric higher is better (↑).

	Precision ↑	Recall ↑	F-Measure ↑	Specificity ↑	Accuracy ↑	Precision ↑	Recall ↑	F-Measure ↑	Specificity ↑	Accuracy ↑
NN	**0.83**	0.29	**0.39**	**0.99**	**0.88**	**0.34**	0.49	**0.29**	**0.63**	**0.60**
NN DA	**0.47**	**0.59**	**0.47**	0.76	0.73	**0.38**	0.39	**0.32**	**0.80**	**0.74**
KISS [17]	0.32	0.30	0.28	**0.82**	**0.74**	0.15	0.93	0.26	0.09	0.21
KISS [17] DA	0.23	**0.59**	**0.31**	0.56	0.56	0.15	0.97	0.26	0.04	0.18
BRSVM [1]	**0.37**	0.27	0.18	0.79	0.70	0.15	**1.00**	0.26	0.00	0.15
BRSVM [1] DA	0.32	0.23	0.17	**0.85**	**0.74**	0.15	**1.00**	0.26	0.00	0.15
RANKSVM [24]	0.00	**0.65**	0.17	0.35	0.38	0.15	**0.98**	0.26	0.03	0.17
RANKSVM [24] DA	0.00	0.36	0.12	0.64	0.58	0.15	**0.98**	0.26	0.03	0.17

Table 3. Re-identification results for Dataset 1: (left) Intra flight, and (right) Inter flight. In each case Precision, Recall and F-measure are averaged across all three flights. Higher is better for all metrics.

ods including KISS, BRSVM and RankSVM. In dramatic contrast to the standard ordering of results obtained in the literature [3, 1, 24], where discriminatively trained models significantly outperform simple nearest-neighbour; our results show that in the MRP context, the simplest NN method is generally best. This is true overall for Dataset 1 with all three tasks, as well as the significantly larger Dataset 2 for the watchlist task. This is due to the intrinsic challenge of MRP re-id that there is no possibility to learn view-specific models.

In order to apply discriminative models to our MRP data, we transferred models trained on VIPER. However, this may not be effective because the MRP video is more variable and unconstrained. Meanwhile, the strong discriminative models have evidently over fitted to the more constrained viewing conditions in VIPER. NN, in contrast, is more reliable because it doesn't train a strong discriminative model and thus cannot over fit in this sense.

(2) Simpler models are better overall The overall ordering of the results is $NN > KISS > BRSVM$. This generally reflects the model complexity, with NN being the simplest. BRSVM being the most complex (due to RBF kernels on concatenated data), and KISS being in between. This ordering also reflects the importance of pairwise training data volume to the model, with KISS and BRSVM both requiring fairly large volumes of training data from the same view in order to perform well.

	First rank ↓	Last rank ↓	Mean rank ↓	Av Prec ↑
KISS (ED)	1.66	64.44	20.79	0.57
KISS-DA (ED)	3.29	60.68	21.40	0.56
KISS	1.25	81.31	25.90	0.53
KISS-DA	3.50	81.65	30.08	0.35

Table 4. Attempting to improve the performance of KISS [17] on the watchlist task by training on all available data (ED). Results are from a single flight in Dataset 1.

(3) Domain adaptation can help – but it helps NN significantly more than discriminative models. Comparing the vanilla condition of each model with the domain adaptation condition (-DA suffix), we see that domain adaptation doesn't make much consistent difference for the watchlist experiment (Table 2), but it sometimes makes a significant difference in the re-identification experiment (Table 3). However, KISS for example is improved from mAP of 0.28 to 0.31 with domain adaptation; while NN is improved much more significantly from mAP of 0.39 to 0.47. That domain-adaptation can help is in one sense not surprising (the MRP video has different statistics to VIPER and aligning the distributions should help), but in another sense surprising (the MRP video is only a *domain* in a very limited sense – because the view varies so much there is hardly a consistent set of statistics $p(X_t)$ to adapt toward). Meanwhile, the fact that it helps NN more than KISS is understandable because KISS still suffers from over fitting to the particular source data (VIPER).

(4) Discriminative models cannot be "fixed" for MRP by adding more conventional training data. The significance of the previous results – with respect to limitations of the discriminative models – could be questioned on the grounds of whether VIPER data is *representative* enough for the variety of views obtained by the MRP. To test this, we re-trained the KISS model using the union of the four largest benchmark re-id datasets to date, including VIPER, CUHK, GRID and PRID, thus greatly increasing the volume and variety of data used. Table 4 compares the watchlist verification results when training KISS only on VIPER versus training on all existing datasets (ED suffix). Clearly using all the extra data makes only a minor difference to the performance.

5.3 Person Count Evaluation

As a final example application, we perform person counting on the flight videos. This is computed as a by-product of open-world re-identification: each identified connected component of the detections defines a distinct person. In general NN and NN-DA provide a near best or best estimate in each case, as seen in Table 5.

	Actual	NN	KISS	BRSVM	NN-DA	KISS-DA	BRSVM-DA	RankSVM
Flight1	6.0	±16.0	±23.0	±79.0	±7.0	±20.0	±37.0	±102.0
Flight2	7.0	±0.0	±0.0	±5.0	±1.0	±3.0	±2.0	±2.0
Flight3	10.0	±40.0	±13.0	±1.0	±6.0	±92.0	±3.0	±27.0
Average	7.7	±18.7	±12.0	±28.3	±4.0	±38.3	±14.0	±42.3
	Actual	NN	KISS	BRSVM	NN-DA	KISS-DA	BRSVM-DA	RankSVM
Flight1⩽2	7.0	±5.0	±0.0	±38.0	±0.0	±0.0	±74.0	±48.0
Flight2⩽3	10.0	±0.0	±13.0	±21.0	±6.0	±5.0	±0.0	±1.0
Flight1⩽3	10.0	±0.0	±6.0	±0.0	±3.0	±7.0	±84.0	±226.0
Average	9.0	±1.7	±6.3	±19.7	±3.0	±4.0	±52.7	±91.7

Table 5. Person counts in Dataset 1. Result for each method is shown as the average error between the estimated and true count. (Lower is better) (upper) **Intra**-flight condition, (lower) **Inter**-flight condition.

6 Discussion

6.1 Summary and Key Results

Based on the experiments and analysis in the previous section, we drew the following conclusions: 1. NN is the best method for MRP re-id, 2. In general simpler methods outperform more complex methods, 3. Unsupervised domain adaptation can improve MRP re-id, 4. The challenge is intrinsic to the nature of benchmark datasets being captured by static cameras, and the MRP dataset being captured by a dynamic camera.

6.2 Implications for future work

Given these insights, we highlight the following implications for future work:

1. Current re-id research has been too focused on learning dataset specific models, leading to dataset bias [30]. Analogous to research trends in more general computer vision [16], developing methods that avoid bias and generalise across datasets is necessary to fully exploit the potential of reid to MRPs.
2. Domain adaptation methods can potentially help adapt re-id methods across scenarios with different data statistics. However while most domain adaptation methods require some supervision in the target domain, it is important that DA methods used in this context are unsupervised, since live annotation of MRP detections is implausible. In the current results, a completely disjoint unsupervised DA module [8] is able to make an impact. Investigating tighter integration of the DA and re-id mechanism is likely to be fruitful.
3. Conventional re-id and DA [8] methods assume the target task is a distinct and discrete context. The continually varying nature of MRP view, and hence data statistics, means that it may be important to treat MRP as an online rather than a discrete adaptation process. This is a somewhat unique aspect of DA for re-id in contrast to more general vision problems [30, 16].

4. Consideration of the MRP task highlights the intrinsically open-world nature of re-id which has largely been ignored for convenience by prior research. In this study we addressed this by a very simple strategy of threshold learning. However, more effort should be put toward developing more systematic and optimal methods to resolve open-world ambiguity.
5. Our new continuously-varying view dataset has a total of 51,922 unconstrained person detections across six flights resulting in hundreds of identities that partially overlap across three outdoor zones. This challenging MRP dataset is qualitatively different to existing re-id datasets, and will help drive the research challenges identified above.

6.3 Potential Applications

Finally, given the partial success obtained so far, we discuss some speculative applications for MRP technology.

Open vs. Closed-loop MRP: Our first re-identification case for MRP is an open-loop scenario where the re-identification task does not directly have any impact on the travel path of the vehicle; but data from the vehicle still enables analysis and detection albeit in a passive sense. In this mode of operation, the MRP will likely either be under control of a human operator, or will follow a set of preconfigured waypoints along a patrol-route, with the video sensor data available for analysis either in near real-time, or after the MRP has returned home. This is conceptually closest to the standard re-identification problem.

In contrast, closed-loop MRP control may be fully or semi-automated and critically, may permit the MRP to automatically adapt a regular patrol-route or journey for optimal performance on specific re-identification tasks. For example, re-id quality-control to move the MRP to get a better view when current re-id is too ambiguous [26]. For a given flight time or length, this then leads into an interesting trade-off between re-id accuracy of each individual versus coverage: the fraction of total people captured in a zone in total [28].

Acknowledgements Ryan Layne is supported by a EPSRC CASE studentship supported by UK MOD SA/SD.

References

1. Avraham, T., Gurvich, I., Lindenbaum, M., Markovitch, S.: Learning implicit transfer for person re-identification. In: Workshop on Re-Identification, ECCV (2012)
2. Avraham, T., Gurvich, I., Lindenbaum, M., Markovitch, S.: Learning Implicit Transfer for Person Re-identification. In: European Conference on Computer Vision, International Workshop on Re-identification. Florence, Italy (2012)
3. Bischof, H., Roth, P.M., Hirzer, M., Wohlhart, P., Kostinger, M.: Large scale metric learning from equivalence constraints. In: IEEE Conference on Computer Vision and Pattern Recognition (2012)
4. Brand, Y., Avraham, T., Lindenbaum, M.: Transitive Re-identification. British Machine Vision Conference (3) (2013)
5. Clarke, R.: Understanding the drone epidemic. Computer Law & Security Review 30(3) (2014)

6. Dollár, P., Appel, R., Belongie, S., Perona, P., Doll, P.: Fast Feature Pyramids for Object Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence (2014)
7. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: IEEE Conference on Computer Vision and Pattern Recognition (2010)
8. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: IEEE Conference on Computer Vision and Pattern Recognition. Ieee (2012)
9. Gong, S., Cristani, M., Loy, C.C., Hospedales, T.M.: Person Re-identification, chap. The Re-Identification Challenge. Springer (2013)
10. Gong, S., Cristani, M., Yan, S., Loy, C.C. (eds.): Person Re-identification. Springer (2014)
11. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: European Conference on Computer Vision. Marseille, France (2008)
12. Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person re-identification by descriptive and discriminative classification. In: SCIA - Proceedings of the 17th Scandinavian conference on Image analysis. SCIA'11 (2011)
13. John, V., Englebienne, G., Krose, B.: Solving person re-identification in non-overlapping camera using efficient gibbs sampling. In: British Machine Vision Conference (2013)
14. Karaman, S., Bagdanov, A.D.: Identity inference: Generalizing person re-identification scenarios. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) Workshop on Re-Identification, ECCV. Lecture Notes in Computer Science, vol. 7583. Springer (2012)
15. Khedhrer, M.I., el Yacoubi, M.A., Dorizzi, B.: Multi-shot surf-based person re-identification via sparse representation. In: Advanced Video Surveillance Systems (2013)
16. Khosla, A., Zhou, T., Malisiewicz, T., Efros, A., Torralba, A.: Undoing the damage of dataset bias. In: European Conference on Computer Vision. Florence, Italy (2012)
17. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. IEEE Conference on Computer Vision and Pattern Recognition (2012)
18. Layne, R., Hospedales, T.M., Gong, S.: Person Re-identification by Attributes. In: British Machine Vision Conference (2012)
19. Layne, R., Hospedales, T.M., Gong, S.: Domain Transfer for Person Re-identification. In: Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams (ARTEMIS). Barcelona, Spain (2013)
20. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning. In: Asian Conference on Computer Vision (2012)
21. Loy, C.C., Xiang, T., Gong, S.: Time-Delayed Correlation Analysis for Multi-Camera Activity Understanding. International Journal of Computer Vision 90(1) (2010)
22. Ma, A.J., Yuen, P.C., Li, J.: Domain Transfer Support Vector Ranking for Person Re-Identification without Target Camera Label Information. In: IEEE International Conference on Computer Vision (2013)
23. Pan, S.J., Yang, Q.: A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering 22(10) (2010)
24. Prosser, B., Zheng, W.S., Gong, S., Xiang, T.: Person Re-Identification by Support Vector Ranking. In: British Machine Vision Conference (2010)
25. Raja, Y., Gong, S.: Person Re-identification, chap. Scalable Multi-Camera Tracking in a Metropolis. Springer (2013)

26. Salvagnini, P., Bazzani, L., Cristani, M., Murino, V.: Person re-identification with a ptz camera: An introductory study. In: IEEE International Conference on Image Processing (2013)
27. Schwartz, W., Davis, L.: Learning discriminative appearance-based models using partial least squares. In: Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on (2009)
28. Sommerlade, E., Reid, I.: Information-theoretic active scene exploration. In: IEEE Conference on Computer Vision and Pattern Recognition (2008)
29. Tarjan, R.: Depth-First Search and Linear Graph Algorithms. SIAM Journal on Computing 1(2) (1972)
30. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: IEEE Conference on Computer Vision and Pattern Recognition (2011)
31. Vezzani, R., Baltieri, D., Cucchiara, R.: People Re-identification in Surveillance and Forensics: a Survey. ACM Computing Surveys 1(1) (2013)
32. Wang, X., Zhao, R.: Person Re-identification, chap. Person Re-identification: System Design and Evaluation Overview. Springer (2013)
33. Zhao, R., Ouyang, W., Wang, X.: Unsupervised Salience Learning for Person Re-identification. In: IEEE International Conference on Computer Vision (2013)