

# Transfer Re-identification: From Person to Set-based Verification

Wei-Shi Zheng<sup>1</sup>, Shaogang Gong<sup>2</sup>, and Tao Xiang<sup>2</sup>

<sup>1</sup>School of Information Science and Technology, Sun Yat-sen University, China

<sup>2</sup>School of Electronic Engineering and Computer Science, Queen Mary University of London, UK

wszheng@ieee.org, sgg@eecs.qmul.ac.uk, txiang@eecs.qmul.ac.uk

## Abstract

*Solving the person re-identification problem has become important for understanding people’s behaviours in a multi-camera network of non-overlapping views. In this work, we address the problem of re-identification from a set-based verification perspective. More specifically, we have a small set of target people on a watch list (a set) and we aim to verify whether a query image of a person is on this watch list. This differs from the existing person re-identification problem in that the probe is verified against a small set of known people but requires much higher degree of verification accuracy with very limited sampling data for each candidate in the set. That is, rather than recognising everybody in the scene, we consider identifying a small set of target people against non-target people when there is only a limited number of target training samples and a large number of unlabelled (unknown) non-target samples available. To this end, we formulate a transfer learning framework for mining discriminant information from non-target people data to solve the watch list set verification problem. Based on the proposed approach, we introduce the concepts of multi-shot and one-shot verifications. We also design new criteria for evaluating the performance of the proposed transfer learning method against the i-LIDS and ETHZ data sets.*

## 1. Introduction

In recent years, matching people across disjoint camera views in a multi-camera system, known as the person re-identification problem [14, 10, 20, 11, 7, 23], has gained increasing interest. If a target disappears from one view in a large area of public space covered by multiple non-overlapping cameras, person re-identification aims to find the same person in another view at a different location/time.

Existing work on person re-identification focuses on finding distinctive feature representation and learning discriminant models. Popular feature representations of people’s appearance include color histogram [14, 10], principal axis histogram [11], rectangle region histogram [4],

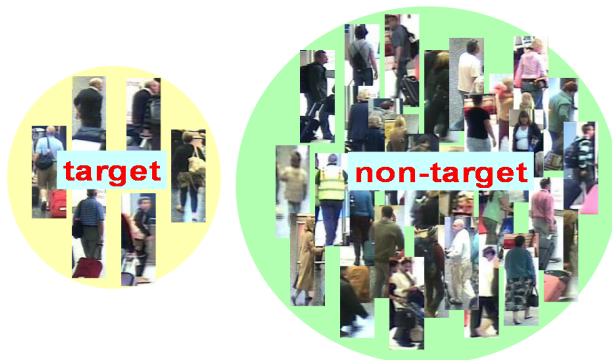


Figure 1. On the left hand side is a small set of target person (watch list) images, and on the right hand side is a large sampling set of non-target person images whose labels are not available.

graph representation [8], spatial co-occurrence representation [20], multiple feature based representation [10, 7]. Due to the large intra-class and inter-class variations of people’s appearance [23], these features may not be reliable under changes in all conditions (e.g. view angle, lighting, occlusion). To address this problem, a number of feature quantification methods have been proposed in the literature including Adaboost [10], Primal RankSVM [16] and Probabilistic Relative Distance Comparison (PRDC) [23].

However, there are two issues that remain unsolved:

- First, person re-identification in existing literature is always treated as a traditional image retrieval or recognition problem. It is assumed that both the gallery set and probe/query set contain the same subjects. However, in a more realistic public environment, the gallery set is much bigger than the probe set and it becomes very difficult to match against everybody in the gallery set exhaustively. As the number of people increases, their re-identification accuracy decreases significantly, especially in unconstrained public spaces [16, 23].
- Second, any person’s appearance is mostly only relatively stable for a short period of time in a real world environment, e.g. a few hours in a day, as people can

dress differently on different days. This is significantly different from other retrieval and classification problems in machine learning such as face recognition. As the current person re-identification problem is defined as a short-period-of-time recognition problem, there are only limited samples available for learning a model capable of re-identifying every person reliably. A learned statistical models for matching two images of a person captured at two different camera views can be easily overfitted.

**Set-based Verification.** We propose a different approach to solving the problem of person re-identification in order to overcome the above hurdles. In particular, we consider a more realistic setting as opposed to the close-world setting used in previous works. Under this setting, person re-identification becomes the problem of verifying a small set of target people, which we call a *watch list*, rather than any person individually in a scene, whilst the available model training samples for the set is very small. More importantly, we also consider the effect of non-target people on diversing such a match. This transforms the person re-identification problem to a problem of verifying each query against the watch list. As the appearance of different people can be rather similar whilst the intra-class variations can be large due to changes in camera view and lighting condition, it is generally hard (if not impossible) for a fully automated system to identify reliably the true targets without the help of human operators. In this context, it is both more realistic and relevant to consider the re-identification problem as matching a small candidate set such that the true target candidates are mostly included there, despite the lack of exact labels of each candidate in the set. To that end, we introduce two set-based person verification problems: one-shot verification and multi-shot verification. One-shot verification is about association to a target person that aims to verify whether a query image is associated with that target person. Multi-shot verification considers the problem of verifying whether a query is within the watch list (set), therefore performing a joint one-shot verification over all target people in the set.

**Our Approach.** For performing verification against non-target people, sampling images of non-target people is important. While there are only limited samples for each target person, one is more likely to have access to a large set of unlabelled person images that are not sourced from the target people. For example, as shown in Figure 1, the left hand side is a set of target people (watch list) and the right side is a sampling set of non-target ones. The use of unlabelled non-target person images aims to extract transferrable discriminant information between the images of target people and those images sampled from the non-targets. To that end, we formulate a set-based transfer learning framework for

verification of each target person and their groupings (in-the-set vs. outside-the-set). Our proposed transfer learning approach is based on the latest bipartite ranking models [16, 23]. The proposed method further explores the useful relative comparison between target and non-target data and makes use of such information to enhance the bipartite ranking analysis between target data, resulting in a more robust statistical model. Our approach is influenced by the general concept of transfer learning which aims to construct more robust statistical learning models that can benefit from shared knowledge between related domains especially when training data is sparse and imbalanced [2, 5, 12, 18, 13, 1, 21].

**Contributions.** The main contributions of this work are: (1) We propose a new and more applicable approach to solve the person re-identification problem by introducing a verification process. The verification process aims to verify a query against a watch list when the visual appearance of the targets are relatively stable. (2) We formulate a novel transfer ranking approach for two types of verification: multi-shot verification and one-shot verification. (3) We propose new criteria for measuring the verification performance of set-based person re-identification.

## 2. The Verification Learning Problem

In the following, we first present the problem before develop the necessary transfer learning techniques.

### 2.1. Problem Formulation

Suppose given  $N_t$  limited target training data from  $m_t$  different target people  $C_1^t, \dots, C_{m_t}^t$  denoted by  $\{\mathbf{z}_i^t, y_i^t\}_{i=1}^{N_t}$ , where  $y_i^t \in \{C_1^t, \dots, C_{m_t}^t\}$  and  $\mathbf{z}_i^t$  denotes the  $i^{th}$  target sample. Also, we are given a big sampling set of unlabelled data from other people denoted by  $\{\mathbf{z}_1^s, \dots, \mathbf{z}_{N_s}^s\}$ , where  $N_s \gg N_t$ . The problem is how to learn a more robust matching model by using these unlabelled non-target data for performing person re-identification on the small set of target people against non-target people.

In some aspect, the above modelling can also be partially translated into separating a target set of people  $C_1^t, \dots, C_{m_t}^t$  from another group of people which are unlabelled. From the multi-task learning point of view, the new modelling of person re-identification here is how to perform the task of group separation learning such that it benefits for the task of re-identification on target people. By combining these two tasks, the verification process for target people against non-target ones is therefore realised.

### 2.2. Mining Transferable Information

Verification is different from recognition. For verification, we are not only aiming to realise the separation between target data of different classes and the separation be-

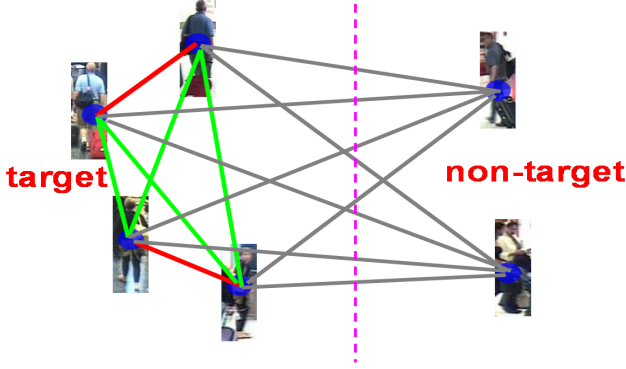


Figure 2. There are three types of variations among a target set and non-target data: 1) the target intra-class variations (red lines); 2) the target inter-class variations (green lines); 3) the inter-class variation between target and non-target images (grey lines).

tween target data and non-target data, but also aiming to infer the difference information between any pair of target data coming from different classes against the one between any pair of target sample and non-target sample, so as to achieve better verification performance. All these three types of information will contribute to our bipartite ranking based verification approach.

More specifically, we wish to learn a score function denoted by  $f(\mathbf{x})$  which is a function of a difference vector  $\mathbf{x}$ , e.g.  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$  or the distance model in Sec. 2.4. Note that the difference vector is computed as an absolute difference vector [23]. For bipartite ranking, the score function for the difference vector  $\mathbf{x}^p$  computed from a relevant pair of images should be larger than the difference vector  $\mathbf{x}^n$  computed from a related irrelevant pair of images, where only one sample for computing  $\mathbf{x}^n$  is one of the two relevant samples for computing  $\mathbf{x}^p$ . The bipartite ranking is then constrained by:

$$f(\mathbf{x}^p) \geq \rho + f(\mathbf{x}^n), \quad (1)$$

where  $\rho$  is a non-negative margin variable.

On target person image data, we wish to infer the intra-class and inter-class variations. To that end, we explore the following target bipartite ranking:

- (1) A score comparison between a pair of relevant target person images (a red line in Figure 2) and a related pair of the irrelevant target person images (a related green line in Figure 2):  $\mathcal{O}_i^t = (\mathbf{x}_i^{t,p}, \mathbf{x}_i^{t,n})$ , where  $\mathbf{x}_i^{t,p}$  is the difference vector [23] computed between a pair of relevant samples of the same target person and  $\mathbf{x}_i^{t,n}$  is the difference vector from a related pair of irrelevant samples, i.e. only one sample for computing  $\mathbf{x}_i^{t,n}$  is one of the two relevant samples for computing  $\mathbf{x}_i^{t,p}$  and the other is a sample from another target person. Denote the set of all these comparison pairs by  $\mathbb{O}_t = \{\mathcal{O}_i^t\}$ .

Then, we wish to realise the following comparison:

$$f(\mathbf{x}_i^{t,p}) \geq \rho + f(\mathbf{x}_i^{t,n}), \text{ for all } \mathcal{O}_i^t \in \mathbb{O}_t. \quad (2)$$

Besides, although the non-target person image data are not labelled, we know they are not from target people. Thus, we can explore the inter-class variations between any pair of target person image and non-target person image, and then the following new comparison can be incorporated for bipartite ranking learning:

- (2) A score comparison between a pair of relevant target person images (a red line in Figure 2) and a related pair of the irrelevant person images between the target person image and any non-target person image (any related grey line in Figure 2):  $\mathcal{O}_i^{ts} = (\mathbf{x}_i^{t,p}, \mathbf{x}_i^{ts,n})$ , where  $\mathbf{x}_i^{t,p}$  is defined in the point (1) and  $\mathbf{x}_i^{ts,n}$  is the difference vector between any sample for computing  $\mathbf{x}_i^{t,p}$  and any un-target person image sample. Denote the set of all these comparison pairs by  $\mathbb{O}_{ts} = \{\mathcal{O}_i^{ts}\}$ . Then, we wish to realise the following comparison:

$$f(\mathbf{x}_i^{t,p}) \geq \rho + f(\mathbf{x}_i^{ts,n}), \text{ for all } \mathcal{O}_i^{ts} \in \mathbb{O}_{ts}. \quad (3)$$

Furthermore, as the appearance of a person can change dramatically across non-overlapping camera views, the objective of this process is to predict a candidate set for assisting human decision making. It is expected that the candidate set contains the target people, illustrated by the query image being recognised as one of the targets on the watch list shown by the solid red line in Figure 3. To that end, we further incorporate the following comparison information.

- (3) A score comparison between a pair of different target images (a green line in Figure 2) and a related pair of irrelevant person images between the target and any non-target (a related grey line in Figure 2):  $\mathcal{O}_i^{tsn} = (\mathbf{x}_i^{t,n}, \mathbf{x}_i^{ts,n})$ , where  $\mathbf{x}_i^{t,n}$  is the difference vector between one of the target images for computing  $\mathbf{x}_i^{t,n}$  and any non-target person image. Denote the set of all these comparison pairs by  $\mathbb{O}_{tsn} = \{\mathcal{O}_i^{tsn}\}$ . Then, we wish to realise the following comparison:

$$f(\mathbf{x}_i^{t,n}) \geq \rho + f(\mathbf{x}_i^{ts,n}), \text{ for all } \mathcal{O}_i^{tsn} \in \mathbb{O}_{tsn}. \quad (4)$$

Minimising the error (risk) function (Eqn. 4) of this comparison ensures that the score of the difference vector computed between a pair of different target person images should be higher than the one computed between a pair of a related target person image and any non-target person image. Such a learning can make target people in the feature space more compact than the non-target ones.

By integrating the above three score comparisons, we develop two transfer bipartite ranking methods, transfer RankSVM and transfer Probabilistic Relative Distance Comparison (PRDC), as follows. They are significantly different from RankSVM and PRDC, because the non-transfer

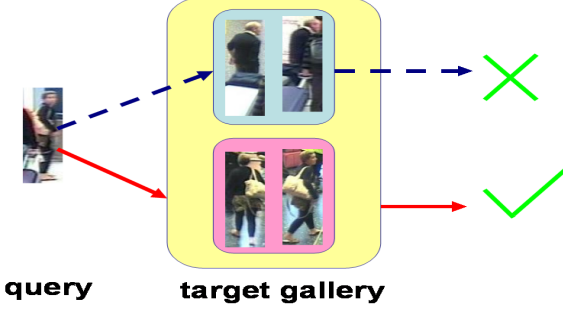


Figure 3. In traditional person re-identification, if the query image is matched to a wrong target person (as shown by the dash blue line), then the output is incorrect. In a generalised set-based person verification by transfer learning, if the query image is matched to one of the target person (as shown by the solid red line), then the output is correct.

models only take the comparison information among target data for bipartite ranking, namely just the point (1) in the above analysis.

### 2.3. Transfer RankSVM

We model the score function  $f(\mathbf{x})$  by  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$  and let  $\rho = 1$  in Sec. 2.2. Noting that by using this score function in Eqns. (2), (3) and (4), we would find that those score comparison will aim to maximise the difference of two scores using a normalised vector  $\frac{\mathbf{w}}{\|\mathbf{w}\|}$  if we divide  $\|\mathbf{w}\|$  on both sides, where the difference is characterised by  $\frac{1}{\|\mathbf{w}\|}$ . Therefore, our transfer RankSVM aims to maximise such a difference via minimising  $\|\mathbf{w}\|$  by formulating the following optimisation problem:

$$\begin{aligned} & \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 \\ & \text{s.t. } \mathbf{w}^T \mathbf{x}_i^{t,p} \geq 1 + \mathbf{w}^T \mathbf{x}_i^{t,n}, \text{ for all } \mathcal{O}_i^t \in \mathcal{O}_t, \\ & \quad \mathbf{w}^T \mathbf{x}_i^{t,p} \geq 1 + \mathbf{w}^T \mathbf{x}_i^{ts,n}, \text{ for all } \mathcal{O}_i^{ts} \in \mathcal{O}_{ts}, \\ & \quad \mathbf{w}^T \mathbf{x}_i^{t,n} \geq 1 + \mathbf{w}^T \mathbf{x}_i^{tsn}, \text{ for all } \mathcal{O}_i^{tsn} \in \mathcal{O}_{tsn}. \end{aligned} \quad (5)$$

Since the feature dimension of person appearance is very high (it is larger than 2700 in this work) and there are a large amount of bipartite ranking, we also develop the corresponding risk based criterion of Eqn.(5) as follows:

$$\begin{aligned} & \min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + \lambda \left[ \frac{1}{|\mathcal{O}_t|} \sum_{\mathcal{O}_i^t \in \mathcal{O}_t} \max\{0, 1 - \mathbf{w}^T (\mathbf{x}_i^{t,p} - \mathbf{x}_i^{t,n})\}^2 \right. \\ & \quad + \frac{\beta}{|\mathcal{O}_{ts}|} \sum_{\mathcal{O}_i^{ts} \in \mathcal{O}_{ts}} \max\{0, 1 - \mathbf{w}^T (\mathbf{x}_i^{t,p} - \mathbf{x}_i^{ts,n})\}^2 \\ & \quad \left. + \frac{\eta}{|\mathcal{O}_{tsn}|} \sum_{\mathcal{O}_i^{tsn} \in \mathcal{O}_{tsn}} \max\{0, 1 - \mathbf{w}^T (\mathbf{x}_i^{t,n} - \mathbf{x}_i^{tsn})\}^2 \right] \end{aligned} \quad (6)$$

where  $\lambda, \beta, \eta \geq 0$ . In the above criterion, we actually in-

tegrate two more parameters  $\beta, \eta$  to balance the impact of knowledge transfer from unlabelled data.

### 2.4. Transfer PRDC

Different from transfer RankSVM which aims to maximise the score difference (i.e. the margin  $\frac{1}{\|\mathbf{w}\|}$ ), transfer PRDC is to further quantify the second-order mutual information between different features and aims to maximise the probability of the realisation of all the comparisons detailed in Sec. 2.2.

Let  $f(\mathbf{x}) = -d(\mathbf{x})$  and  $\rho = 0$  in Sec. 2.2, where  $d(\mathbf{x})$  is a distance function defined by  $d(\mathbf{x}) = \mathbf{x}^T \mathbf{W} \mathbf{W}^T \mathbf{x}$  for some matrix  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_\ell]$ . Then transfer PRDC learns this score function by maximising the following confidence function:

$$\begin{aligned} & \left[ \prod_{\mathcal{O}_i^t \in \mathcal{O}_t} P(d(\mathbf{x}_i^{t,p}) < d(\mathbf{x}_i^{t,n})) \right]^{\frac{1}{|\mathcal{O}_t|}} \\ & \times \left[ \prod_{\mathcal{O}_i^{ts} \in \mathcal{O}_{ts}} P(d(\mathbf{x}_i^{t,p}) < d(\mathbf{x}_i^{ts,n})) \right]^{\frac{\beta}{|\mathcal{O}_{ts}|}} \\ & \times \left[ \prod_{\mathcal{O}_i^{tsn} \in \mathcal{O}_{tsn}} P(d(\mathbf{x}_i^{t,n}) < d(\mathbf{x}_i^{tsn})) \right]^{\frac{\eta}{|\mathcal{O}_{tsn}|}} \end{aligned} \quad (7)$$

where

$$P(d(\mathbf{x}) < d(\mathbf{x}')) = (1 + \exp \{d(\mathbf{x}) - d(\mathbf{x}')\})^{-1}. \quad (8)$$

Maximising this confidence model is equivalent to the following optimisation problem:

$$\begin{aligned} & \mathbf{W} = \arg \min_{\mathbf{W}} r(\mathbf{W}), \text{ s.t. } \mathbf{w}_i^T \mathbf{w}_j = 0, \forall i \neq j \\ & r(\mathbf{W}) = \frac{1}{|\mathcal{O}_t|} \sum_{\mathcal{O}_i^t \in \mathcal{O}_t} \log(1 + \exp \{ \|\mathbf{W}^T \mathbf{x}_i^{t,p}\|^2 - \|\mathbf{W}^T \mathbf{x}_i^{t,n}\|^2 \}) \\ & \quad + \frac{\beta}{|\mathcal{O}_{ts}|} \sum_{\mathcal{O}_i^{ts} \in \mathcal{O}_{ts}} \log(1 + \exp \{ \|\mathbf{W}^T \mathbf{x}_i^{t,p}\|^2 - \|\mathbf{W}^T \mathbf{x}_i^{ts,n}\|^2 \}) \\ & \quad + \frac{\eta}{|\mathcal{O}_{tsn}|} \sum_{\mathcal{O}_i^{tsn} \in \mathcal{O}_{tsn}} \log(1 + \exp \{ \|\mathbf{W}^T \mathbf{x}_i^{t,n}\|^2 - \|\mathbf{W}^T \mathbf{x}_i^{tsn}\|^2 \}). \end{aligned} \quad (9)$$

We call the above model as Transfer PRDC.

### 2.5. Optimisation Algorithm

Due to the limit of space, the optimisation algorithms of the proposed transfer RankSVM and transfer PRDC are not detailed here. Gradient based algorithms for transfer RankSVM and transfer PRDC, which are able to handle the large-scale computation, can be developed similarly as suggested in [3, 23], respectively.

### 3. Multi-shot and One-shot Verifications

After learning the transfer models in Sec. 2.3 and 2.4 where the effect of non-target people has been considered, we can perform verification for person re-identification. The two following types of verification are considered.

**Multi-shot Verification.** Given a query person image  $I$ , the multi-shot verification aims to verify whether this image comes from one of the target person on the watch list. That is multi-shot verification performs a one-to-set verifying.

**One-shot Verification.** Given a query person image  $I$ , the one-shot verification aims to verify whether this image comes from target person  $C_k^t$  and does not come from the others (including the other target people).

The difference between these two verifications is that multi-shot verification tells whether the detected person is within our interest but does not perform verification on the person identity of any query image. The one-shot verification performs the latter but would not be able to measure explicitly the probability that the person of the query image is on the watch list. Their relation is similar to the relation between joint probability density function and marginal probability density function.

## 4. Experiments

### 4.1. Settings

**Datasets.** We use both the i-LIDS Multiple-Camera Tracking Scenario (MCTS) dataset [22, 19, 23] and the ETHZ dataset [17, 6] for evaluation. The i-LIDS MCTS dataset consists of 119 people with a total 476 person images with an average of 4 images, which are captured by multiple non-overlapping cameras indoor at a busy airport arrival hall. Many of these images undergo large illumination change and are subject to occlusions. The ETHZ dataset consists of 146 people and 8555 image in total, where images of person were taken from a moving camera in a busy street scene. Please note that, the labels of all non-target data used in our experiment are assumed to be unknown. The popular VIPeR dataset [9] was not used here because there is only two samples for each person. In our experiments, only one training image for each person (where the other one is for testing) is not enough for implementing RankSVM and PRDC for comparison.

**Competitors & Parameter Settings.** We compare our proposed transfer RankSVM and transfer PRDC with RankSVM [16] and PRDC [23] which have been shown to be more effective than any other existing person re-identification methods. Through all experiments, we fixed the parameters  $\beta, \eta$  to be 0.5 in Eqns. (9) and (6) for the two proposed transfer learning techniques. We also fix the  $\lambda$  in both RankSVM and Transfer RankSVM to be 0.005. This can in general be learned through cross-validation [23]. However, the value in our experiment is more difficult to be determined by cross-validation as we are considering person re-identification on a small set of target people with a handful of training samples, against other unlabelled non-target people.

**Feature Representation.** The popular histogram based feature representation for person re-identification is adopted, which is a mixture of colour features (including RGB, Y-CbCr, HSV color) and texture feature pattern (extracted by Schmid and Gabor filters) [10, 15, 22, 23]. Each person image is represented by a feature vector in a 2784 dimensional feature space.

### 4.2. Verification Task and Performance Metrics

The objective here is to verify whether a query person image comes from the people on the watch list with the presence of non-target person images during the verification. More specifically, for each dataset, we randomly selected all images of  $p$  people (classes) to set up the target data set and the rest to constitute the non-target data set. Then, for the target data set, we randomly divided it into a training set and a testing set, where  $q$  images of each person were randomly selected for training. We also randomly divided the non-target data set into training and testing sets. Such a random division is done by person; that is the images of half of the non-target people in the data set were used as training non-target person images and all the rest images were selected as testing non-target images so that there is no overlap of people between training non-target images and testing non-target images. It is important to note that we assume the labels of these non-target data are unknown. The experiment was conducted 10 times and the average verification rates were then computed. Through the experiment, we tested two scenarios by setting the number of target people to 6 and 10 (i.e.  $p = 6$  or  $p = 10$ ) respectively, and randomly select two images (i.e.  $q = 2$ ) as training samples for each target person.

Since a lot images of non-target people were mixed with the target ones during the verification, we need to quantify the performance how well a true target has been verified, how bad a false target has passed through the verification and their relations. Therefore, we introduce the true target rate (TTR) and false target rate (FTR) as follows:

$$\text{True Target Recognition(TTR)} = \frac{\#TTQ}{\#TQ}, \quad (10)$$

$$\text{False Target Recognition(FTR)} = \frac{\#FNTTQ}{\#NTQ} \quad (11)$$

where

$TQ = \{\text{query target images from target people}\};$

$NTQ = \{\text{query non-target images from non-target people}\};$

$TTQ = \{\text{query target images that}$

are verified as one of the target people}\};

$FNTQ = \{\text{query non-target images that}$

are verified as one of the target people}\}.

Note that for performing one-shot verification for each target person (see Sec. 3), the above metrics can still be used, and in this case the non-target people mean any other person

Methods	i-LIDS		ETHZ	
	$FTR = 30\%$	$FTR = 50\%$	$FTR = 30\%$	$FTR = 50\%$
RankSVM	26.87	50.32	25.82	45.15
Transfer RankSVM	53.25	71.08	63.67	75.22
PRDC	57.27	74.50	64.87	77.69
Transfer PRDC	<b>68.89</b>	<b>81.68</b>	<b>76.08</b>	<b>85.5</b>

Table 1. Multi-shot Verification on 6 People ( $p = 6$ ): True target rate (%) against False target rate for one-shot learning

Methods	i-LIDS		ETHZ	
	$FTR = 30\%$	$FTR = 50\%$	$FTR = 30\%$	$FTR = 50\%$
RankSVM	27.18	49.63	25.68	41.51
Transfer RankSVM	54.03	67.85	63.19	77.24
PRDC	45.38	64.32	64.63	77.12
Transfer PRDC	<b>62.67</b>	<b>75.96</b>	<b>74.19</b>	<b>85.09</b>

Table 2. Multi-shot Verification on 10 People ( $p = 10$ ): True target rate (%) against False target rate on i-LIDS and ETHZ

Methods	i-LIDS		ETHZ	
	$FTR = 30\%$	$FTR = 50\%$	$FTR = 30\%$	$FTR = 50\%$
RankSVM	10.65	17.75	4.59	7.4
Transfer RankSVM	<b>84.33</b>	<b>92.37</b>	<b>91.27</b>	<b>95.34</b>
PRDC	71.53	86.78	88.45	92.77
Transfer PRDC	74.65	86.06	85.95	90.74

Table 3. One-shot Verification on 6 People ( $p = 6$ ): True target rate (%) against False target rate on i-LIDS and ETHZ

except that target person.

In our experiments, the score function  $f(\mathbf{x})$  which is specified in Sec. 2 is used to determine the rank of matching. A value  $s$  is used to threshold these scores and therefore a curve depicting the TTR value against FTR value is reported for each method by changing the threshold value  $s$ . We will also report the TTR when the FTR is fixed. This is similar to the ROC curve in face verification, but it differs in that we are also caring about the verification on whether the query image is belonging to one of the target people (i.e. multi-shot verification), because person re-identification could rely much more on the processing of human operators in real scenario than face recognition due to large intra- and inter- class variations.

### 4.3. Verification Results

**Transfer vs. Non-Transfer.** We compare non-transfer person re-identification with transfer person re-identification on multi-shot verification and one-shot verification, i.e. RankSVM vs. Transfer RankSVM and PRDC vs. Transfer PRDC.

For multi-shot verification as shown in Tables 1 and 2 and Figure 4, both transfer PRDC and transfer RankSVM have achieved significant better performance than their non-transfer versions PRDC and RankSVM, respectively. For example, at least about 10 true mating rate higher for transfer PRDC against PRDC when  $FTR = 0.3$  on i-LIDS and ETHZ, and more than about 25 true mating rate higher for Transfer RankSVM against RankSVM when  $FTR = 0.3$  on

i-LIDS and ETHZ. For one-shot verification as shown in Tables 3 and 4 and Figure 5, Transfer RankSVM consistently outperforms RankSVM and Transfer PRDC still performs better than PRDC on i-LIDS, while slightly negative transfer has been observed for Transfer PRDC on ETHZ (i.e. slightly lower performance).

Nevertheless, from the multi-shot verification to one-shot verification, the transfer learning models perform much more stably and reliably than the non-transfer ones. Our results, especially the comparison between Transfer RankSVM and RankSVM, also show that without learning from unlabelled data, the bipartite ranking may largely fail for person re-identification where un-target instances are involved during the verification.

**Transfer PRDC vs. Transfer RankSVM.** We compare the two proposed transfer learning algorithms here for verification. As shown in Tables 1 and 2 and Figure 4, for multi-shot verification transfer PRDC consistently outperform transfer RankSVM with a large margin (e.g. more than 15% increase in true matching rate and 12% in true matching rate at  $FTR=0.3$  on i-LIDS and ETHZ respectively). However, for one-shot verification as shown in Tables 3 and 4 and Figure 5, transfer RankSVM performs better than transfer PRDC. This may suggest transfer PRDC is more suitable for a joint verification, whilst transfer RankSVM, which explicitly maximises the marginal information, is more suitable for a one-to-one verification.



Methods	i-LIDS		ETHZ	
	$FTR = 30\%$	$FTR = 50\%$	$FTR = 30\%$	$FTR = 50\%$
RankSVM	6.9	16.86	3.7	6.3
Transfer RankSVM	<b>82.76</b>	<b>92.78</b>	<b>92.33</b>	<b>96.05</b>
PRDC	71.37	85.79	88.52	93.80
Transfer PRDC	76.64	87.08	87.16	93.78

Table 4. One-shot Verification on 10 People ( $p = 10$ ): True target rate (%) against False target rate

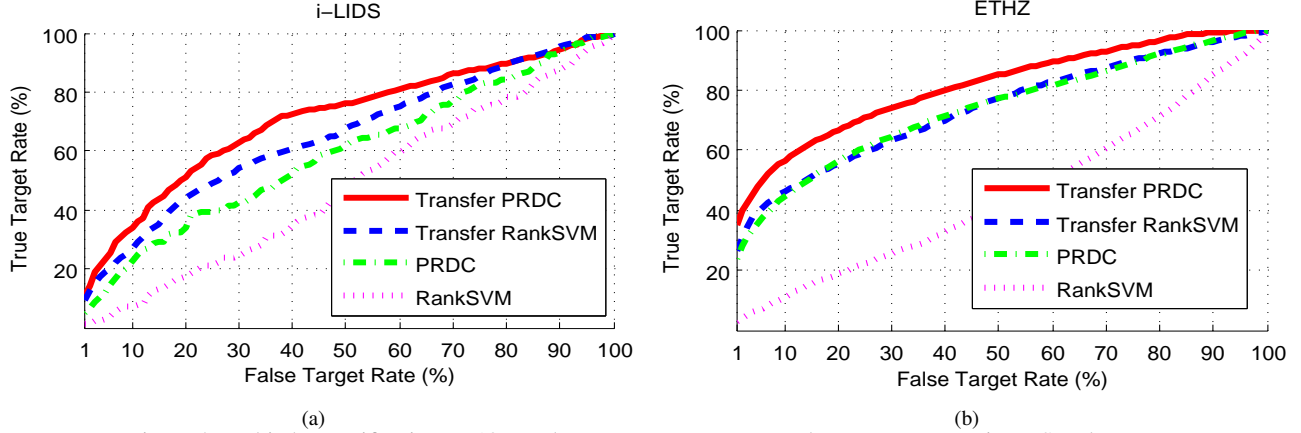


Figure 4. Multi-shot Verification on 10 people: True Target Rate vs. False Target Rate on i-LIDS and ETHZ

Methods	$\beta = \eta$				
	0.1	0.3	0.5	0.7	0.9
Multi-shot verification					
Transfer RankSVM	44.50	49.67	53.25	56.69	57.53
Transfer PRDC	69.31	68.25	68.89	68.02	68.32
One-shot verification					
Transfer RankSVM	84.40	84.23	84.33	84.05	83.50
Transfer PRDC	76.06	75.32	74.65	75.02	74.87

Table 5. True target rate (%) at  $FTR = 0.3$ : Parameter Validation on i-LIDS ( $p = 6$ ).

**Multi-shot Verification vs. One-shot Verification.** It is clear that the one-shot verification always achieved much higher true target rate as compared to the one for multi-shot verification. However, as analysed in Sec. 3, the one-shot verification is a one-to-one verification and cannot make a joint verification explicitly to say whether the person of a query image is one of the several people needed to be watched. It is because one-shot verification still remains the uncertainty whether the query is also any of the other target people on the watch list. The multi-shot verification is therefore to make a prediction by joining all these uncertainty. From another point of view, non-target people may also have better chances to get access in the multi-shot verification case. Thus the relative lower true target rate will be observed for multi-shot verification. It is like the case that the value of a joint probability density function is lower than the one of each marginal probability density function. In addition, as shown above, different transfer learning methods would be suitable for different types of verification.

**Stability of Parameters.** Note that in preceding experiments, we have fixed the parameters  $\beta, \eta$  to 0.5 in Eqns. (9)

and (6) for both proposed models. We finally evaluate the stability of this setting in Table 5. Due to the limit of space, this validation was only conducted on i-LIDS over 6 target people (i.e.  $p = 6$ ). As shown, both transfer PRDC and transfer RankSVM perform stably from  $\beta = \eta = 0.1$  to  $\beta = \eta = 0.9$ , especially around 0.5.

## 5. Conclusion

We have redefined the person-reidentification problem as a set-based verification problem to reflect a more realistic real-world application requirement. To the best of our knowledge, it is the first attempt on addressing the person re-identification problem from this new perspective. Two transfer person re-identification techniques namely transfer PRDC and transfer RankSVM have been developed to solve this problem in the context of transfer learning. Based on these two models, two types of verification for person re-identification have been introduced, namely multi-shot verification and one-shot verification. Our results indicate that mining useful discriminant knowledge from unlabelled non-target people is very useful for performing verification on a small set of target people who are on the watch list.

## Acknowledgment

W.-S. Zheng is supported by Natural Science Foundation Of China (No. 61102111), Specialized Research Fund for the Doctoral Program of Higher Education (20110171120051), the 985 project in Sun Yat-sen Uni-

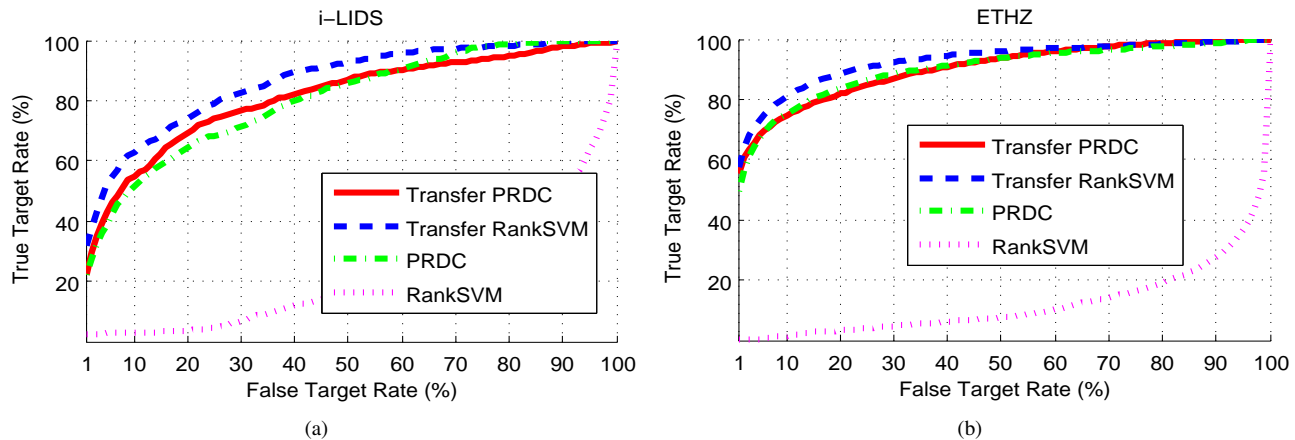


Figure 5. One-shot Verification on 10 people: True Target Rate vs. False Target Rate on i-LIDS and ETHZ

versity with grant no. 35000-3181305 and the NSFC-Guangdong (U0835005). S. Gong and T. Xiang are supported by the UK Home Office CONTEST Programme and Vision Semantics Limited.

## References

- [1] R. Ando and T. Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853, 2005.
- [2] E. Bart and S. Ullman. Cross-generalization: Learning novel classes from a single example by feature replacement. In *Computer Vision and Pattern Recognition*, 2005.
- [3] O. Chapelle and S. S. Keerthi. Efficient algorithms for ranking with svms. *Information Retrieval*, 13:201–215, 2010.
- [4] P. Dollar, Z. Tu, H. Tao, and S. Belongie. Feature mining for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [5] L. Duan, I. W. Tsang, D. Xu, and S. J. Maybank. Domain transfer svm for video concept detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [6] A. Ess, B. Leibe, and L. Van Gool. Depth and appearance for mobile scene analysis. In *IEEE International Conference on Computer Vision*, 2007.
- [7] M. Farenzena, L. Bazzani, A. Perina, M. Cristani, and V. Murino. Person re-identification by symmetry-driven accumulation of local features. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [8] N. Gheissari, T. Sebastian, and R. Hartley. Person reidentification using spatiotemporal appearance. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.
- [9] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE International workshop on performance evaluation of tracking and surveillance*, 2007.
- [10] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision*, 2008.
- [11] W. Hu, M. Hu, X. Zhou, J. Lou, T. Tan, and S. Maybank. Principal axis-based correspondence between multiple cameras for people tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):663–671, 2006.
- [12] F. F. Li, R. Fergus, and P. Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):594–611, 2006.
- [13] S. J. Pan, J. T. Kwok, and Q. Yang. Transfer learning via dimensionality reduction. In *AAAI Conference on Artificial Intelligence*, pages 677–682, 2008.
- [14] U. Park, A. Jain, I. Kitahara, K. Kogure, and N. Hagita. Vise: Visual search engine using multiple networked cameras. In *International Conference on Pattern Recognition*, 2006.
- [15] B. Prosser, S. Gong, and T. Xiang. Multi-camera matching using bi-directional cumulative brightness transfer functions. In *British Machine Vision Conference*, 2008.
- [16] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *British Machine Vision Conference*, 2010.
- [17] W. Schwartz and L. Davis. Learning discriminative appearance based models using partial least squares. In *Brazilian Symposium on Computer Graphics and Image Processing*, pages 322–329, 2009.
- [18] A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In *Computer Vision and Pattern Recognition*, 2004.
- [19] UK. Home Office i-LIDS multiple camera tracking scenario definition. 2008.
- [20] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu. Shape and appearance context modeling. In *IEEE International Conference on Computer Vision*, 2007.
- [21] Y. Zhang and D. Yeung. A convex formulation for learning task relationships in multi-task learning. In *Conference on Uncertainty in AI*, 2010.
- [22] W.-S. Zheng, S. Gong, and T. Xiang. Associating groups of people. In *British Machine Vision Conference*, 2009.
- [23] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.