Partial Person Re-identification

Wei-Shi Zheng[1], Xiang Li[1], Tao Xiang[2], Shengcai Liao[3], Jianhuang Lai[1], and Shaogang Gong [2]
[1]School of Information Science and Technology, Sun Yat-sen University, China
[2]School of Electronic Engineering and Computer Science , Queen Mary University of London, UK
[3]NLPR, Institute of Automation, Chinese Academy of Sciences, China.
wszheng@ieee.org, lixiang651@gmail.com, t.xiang@qmul.ac.uk, scliao@nlpr.ia.ac.cn
stsljh@mail.sysu.edu.cn, s.gong@qmul.ac.uk

Abstract

We address a new partial person re-identification (re-id) *problem, where only a partial observation of a person is available for matching across different non-overlapping camera views. This differs significantly from the conventional person re-id setting where it is assumed that the full body of a person is detected and aligned. To solve this more challenging and realistic re-id problem without the implicit assumption of manual body-parts alignment, we propose a matching framework consisting of 1) a local patch-level matching model based on a novel sparse representation classification formulation with explicit patch ambiguity modelling, and 2) a global part-based matching model providing complementary spatial layout information. Our framework is evaluated on a new partial person re-id dataset as well as two existing datasets modified to include partial person images. The results show that the proposed method outperforms significantly existing re-id methods as well as other partial visual matching methods.*

1. Introduction

Person re-identification (re-id) has been studied extensively in the past five years [5, 30, 4, 37, 10, 20, 36]. It aims to re-identify a target person in a new view after he/she disappears from another view in a large public space covered by multiple non-overlapping (disjoint) cameras. Solving the re-id problem has many applications in video surveillance for public security and safety.

To match a person across views, one has to deal with large appearance changes of the person caused by a variety of condition changes including lighting, view angle, pose, and occlusion. Among these challenges, occlusion is one of the hardest to tackle because the information loss is irreversible. Occlusions are commonplace in a crowded public space with background clutters, where they can be caused by other people in the scene, or static obstacles such as walls/pillars. Sometimes occlusions can be caused deliberately. For example, when a crime has been reported, the police or CCTV operator who examines the surveillance footage often discovers that only partial body of the suspect is visible at the crime scene because one deliberately tried to hide one's appearance (see Fig. 1). The goal is to find the same person in full body appearance in other camera views given only a partial probe image. We call this the *partial person re-identification* problem.

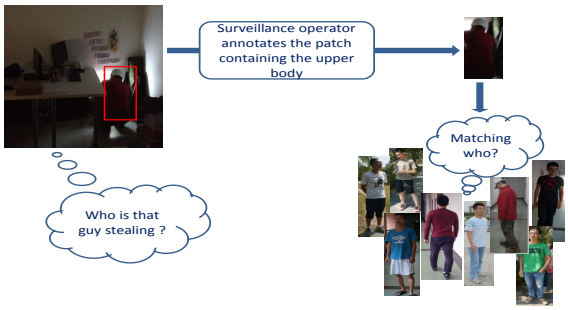


Figure 1. An illustration of the partial person re-id problem. Here, an operator wishes to know who the person is stealing in an office. The only visible cue is the upper body clothing. The operator may crop manually the visible parts and uses an automated re-id system to match against a set of observations from elsewhere.

There are two computational challenges for solving the partial person re-id problem (see examples in Fig. 2). First, with only partial information it is less discriminative thus more likely to be mismatched to the wrong person. For example if only a pair of blue jeans are visible, it is difficult to use them to distinguish people as so many other people may also wear similar jeans, whilst other more distinct body parts were occluded. Second, it is difficult to determine against which part of the full body the partial observation should be matched. The partial appearance can be a random part of a person and which body part it belongs to can be ambiguous. For example if only a small part of a clothing

Figure 2. Examples of partial person images (first row), and the input partial part annotated by an operator for recognition (second row) and the corresponding non-partial images (third row) in the new Partial REID dataset introduced in this paper.

article is observed, it is extremely hard to tell whether it is part of the upper or bottom garment. Manual part alignment is a solution but it is unscalable. A perfect body part detector may also solve the problem but such a detector does not exist under severe occlusions.

Existing person re-id approaches cannot solve the partial person re-id problem. Regardless whether they are designed for computing either cross-view invariant features or distance metrics [4, 5, 30, 10, 20, 38, 14, 9, 26, 42, 22, 27, 25, 15, 37, 36, 33, 21, 18], the existing models all assume that a full body appearance of each person is available. This is also reflected in the existing benchmarking re-id datasets, most of which consist of no occlusion or negligible occlusions with carefully (manually) cropped full body images. None of them is designed to test/evaluate re-id given severe partial observations as probe images without manual alignment against fully observed images in a gallery.

To address the partial person re-id problem without manual alignment of partial observation, we formulate a new re-id framework consisting of two matching components. The first component is based on patch-level local-to-local matching. Specifically, we decompose the probe partial image and the gallery images into small local patches. To tackle the ambiguous nature of patch appearance during the patch-based matching, we introduce a novel Ambiguity-sensitive Matching Classifier (AMC) which computes an ambiguity score at the patch-level between a probe and each gallery patch. In order to perform gallery patch selection during the matching, AMC is formulated in the context of sparse representation classification. Specifically, the ambiguity scores are used in the sparse modelling to guide a selective search of similar gallery patches for each probe patch in order to reduce the risk of mismatch. The second component of our model is based on a global-to-local matching by a detection-based matching model that treats the visible part as an object and slides it exhaustively over a gallery person image in order to provide the complementary matching information to the local-patch based model. Finally, the outputs of the two complementary components are fused in order to make the whole framework more robust against occlusion and background clutter.

The main contributions of this work are: (1) it is the first work that defines the partial matching problem for person re-id, and (2) it proposes an effective computer vision model to fully address this problem. Extensive experiments are conducted for evaluation. In particular, since the existing benchmarks largely ignored this partial re-id problem, we contribute a new dataset called Partial REID dataset, which is specifically designed for this problem with a great deal more partial instances (see Fig. 2). Moreover, modification on two existing datasets are also carried out to simulate the partial re-id problem. Experimental results on all three datasets show that the proposed model is effective in addressing the partial person re-id problem, achieving significant improvement over representative re-id methods applied to the same problem (at least 12% increase on Rank 1 matching accuracy). It also outperforms a number of alternative models designed for other partial matching problems such as the less challenging partial face recognition.

2. Related Works

Person re-id remains an unsolved problem due to large intra-class and inter-class variations caused by lighting change, pose/view change, and (self-)occlusion. To address these challenges, most of recent works can be categorised into two groups: methods that extract invariant and discriminant features [5, 30, 4, 10, 20, 38, 14] and methods that learn robust metrics or subspace for matching [5, 9, 26, 42, 22, 27, 25, 15, 37, 36, 33, 21, 18]. There are also works on the generalisation of re-id, e.g., transfer-based [13], post-rank based [19], watch-list based for the more realistic open-set setting [41], and spatial-temporal based [16] re-id methods. Recently, deep learning [14] and video-based modelling [29] are also introduced for person re-id. However, all these works either explicitly or implicitly assume matching of full-body appearances with negligible missing parts between the probe and gallery images. They are thus ineffective for solving the partial re-id problem as shown in our experiments.

A number of existing person re-id methods consider a part-based model which offers a partial solution to the occlusion problem in re-id. A pictorial model was employed for part-to-part matching for person re-id in [3]. Xu et al. [34] introduced a cluster sampling based compositional part-based template method. However, these models rely on prior knowledge about the part-based templates. In a practical scenario, the observed part of a person may not be a regular part defined by the templates. Lian et al. [16] introduced a spatial-temporal Bayesian model which is able to handle occlusions caused by multiple people walking together and Zheng et al. [39] proposed group context to

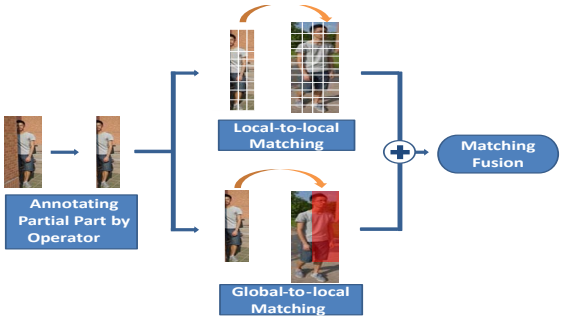


Figure 3. Our partial person re-identification framework.

overcome self-occlusion. 3D model [2] is also considered. However, these two methods still assume that the full body of a person is detected (manually cropped). Under severe occlusions, such full body detection is not obtainable even manually. Although sparse model [8] is used for solving the occlusion problem, it assumes that the alignment is given.

Beyond person re-id, occlusion has been studied extensively in other computer vision problems. It is an especially important topic in face recognition, since faces are often occluded or self-occluded [12]. Recently, sparse representation or dictionary learning has been utilised for solving the occlusion problem in face recognition [32, 43, 6, 35]. Liao et al. [17] proposed a multi-task sparse representation for solving the partial face recognition problem. In order to further take the structure of occlusion as prior knowledge into consideration, Min et al. [23] proposed to first detect the occlusion parts. Meng and Zhang [35] proposed to use an occlusion dictionary to describe the occlusion, and further improvements were reported in [24, 1]. Weng et al. [31] proposed a robust feature matching method and Hu et al. [7] proposed an instance-to-class metric for partial face recognition. Our method is related to them in that our local patch matching model is also based on a sparse representation framework. However, the partial person re-id problem has two unique characteristics that distinguish it from the occlusion modelling in these methods for face recognition. (1) The occlusion problem in person re-id is much more unstructured since a person's body appearance (e.g. clothing) and deformation are much more diverse than those of face. (2) The observed part of a person is also much more diverse, with different sizes and resolutions, and thus more ambiguous (see Fig. 2), compared with partial face, since face has more uniform structure. Due to the above two reasons, the occlusion types on a body are more varied, not only in texture but also in the resolution. Consequently it is much harder to detect what is the occluded part and also construct an occlusion dictionary that describe all the possible occlusion cases on body.

To overcome these unique challenges in person re-id, in this work, we propose an Ambiguity-sensitive Matching Classifier (AMC) in order to model explicitly the diversity and ambiguity of the occlusion patterns in partial person re-id in our new sparse representation classification formulation. Deploying the AMC to compute an ambiguity score enables our model to learn a more robust sparse representation against measurement noise and intrinsic appearance ambiguity. Unable to cope with these person re-id specific challenges, some existing models for handling the occlusion in face recognition such as [23, 35, 24, 1] cannot be directly applied for the partial person re-id problem. Others yield inferior performance compared to our model as demonstrated in our experiments. Note that although the local modelling adopted in [28] is related to the ambiguity score modelling in our proposed AMC model, it does not result in sparse coding. In addition, the method of [28] requires to compute the inverse of a dictionary size matrix, which becomes too costly for a large scale patch-based dictionary (e.g., larger than 10^4), which is typical in re-id.

3. Methodology

In this work, we assume that in partial person re-id a probe image contains only partial body of a person and the task is to match this partial observation with a gallery consisting of full-body images. This is based on a practical scenario where a human operator has manually cropped the observed body part and sent a query to a re-id system to search for the same person in another camera view with full-body images. Note that we do not assume the body part has been named/labelled in the probe image, e.g. legs, left part of a torso. This is challenging even for human operators.

As shown in Fig. 3, our partial person re-id framework has two main matching components: a local-to-local re-id model and global-to-local re-id model. Using the local-to-local matching model, we decompose the partial observation into small patches, and perform matching at the patch level. In contrast, using the global-to-local matching model, we take the partial observation as a whole and search it in each gallery image using a sliding window search strategy. Both models have clear pros and cons: Local patch is less affected by view/pose changes and non-rigid deformations of human body. However, it contains less information than the whole part, and the spatial layout information of different patches is ignored during matching, thus incurring the mis-alignment problem. In contrast, using the whole partial observation as a searching unit enforces spatial layout consistency, but suffers greatly from the view/pose changes and body deformations. Therefore in our framework, the two models are combined to produce the final matching model.

3.1. Local-to-Local Matching

The local-to-local matching model is based on pairwise patch-based matching by sparse coding. After decomposing both gallery and probe images into regular grid patches, we first construct multiple sets of patch level descriptors

from gallery images to form a dictionary. More specifically, multi-patch features are first extracted for each image. Suppose k_c patches, $p_{c_1}, p_{c_2}, \cdots, p_{c_{k_c}}$, are obtained for class (person) c in the gallery. If class c has multiple person images, we collect the patches from all of them. Next, for each patch p_{c_i}, we generate a M-dimensional feature vector $\mathbf{d}_{c_i}$ for representation (see Sec. 4.1 for details). Then, we have k_c patch feature vectors $\mathbf{d}_{c_1}, \mathbf{d}_{c_2}, \cdots, \mathbf{d}_{c_{k_c}}$. Let

$$\mathbf{D}_c = [\mathbf{d}_{c_1}, \mathbf{d}_{c_2}, \cdots, \mathbf{d}_{c_{k_c}}] \,. \tag{1}$$

So the gallery dictionary is built as

$$\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_C] \,. \tag{2}$$

$\mathbf{D}$ has a total of $K = \sum_{c=1}^{C} k_c$ patches, resulting in an $M \times K$ dictionary, where C is the number of classes in the gallery.

Ambiguity-sensitive Matching. In order to achieve the patch-based matching without manual alignment, we wish to compute an ambiguity-sensitive coding of each probe patch feature $\mathbf{y}_i$ with respect to the gallery dictionary $\mathbf{D}$, and we denote such a coding as $\mathbf{x}_i$. We can estimate $\mathbf{x}_i$ by minimising a reconstruction error using the dictionary:

$$\min_{\mathbf{x}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2, \tag{3}$$

That is, we aim to associate the probe partial image with a person class label from the gallery images that contributes the highest number of dictionary patches selected for the reconstruction of the probe. However, there is a flaw in the above reconstruction process – for each patch of the probe image, as the dictionary consists of a pool of all the patches from all the gallery images, it is free to use any patch from any class/person. In other words, for reconstructing each probe patch, there is no constraint on the selection of gallery patches such that those patches similar to the probe are more likely to be selected for the reconstruction. As a result, some visually distinct patches can and will be selected because a linear combination could yield the best reconstruction. Critically, this cannot be addressed explicitly by a sparsity constraint on the coding $\mathbf{x}_i$ (introduced later). This is essentially an unsolvable problem, because the label of the probe is assumed unknown. However, it is still possible to alleviate this problem by avoiding the selection of a gallery patch that looks nothing like the probe patch, but when combined with other patches, produces the minimal reconstruction error. To this end, we measure the visual dissimilarity as matching ambiguity at the patch level, between the probe patch and each of the gallery patch, so as to constrain the gallery patch selection for reconstruction.

More specifically, given a probe patch, each gallery patch's suitability for reconstructing the probe patch is measured using a score, which we call *ambiguity* score. The ambiguity score is defined by a Gaussian-shaped kernel $G_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{x^2}{2\sigma^2})$, where σ^2 is estimated by averaging the Euclidean distance between patches in the training images. Now, given a probe patch feature $\mathbf{y}_i$, for every patch feature $\mathbf{d}_j$ from the gallery dictionary ($\mathbf{D} \in \mathbf{R}^{M \times K}$), the *ambiguity* score is denoted as

$$pa_{i_j} = -G_\sigma(D(\mathbf{y}_i, \mathbf{d}_j)) \,, j = 1, 2, \cdots, K \,, \tag{4}$$

where $D(\cdot)$ is Euclidean distance. For a probe image with n patches, a patch ambiguity score matrix is computed:

$$\mathbf{pa} = [\mathbf{pa}_1, \mathbf{pa}_2, \cdots, \mathbf{pa}_n] \,, \tag{5}$$

where $\mathbf{pa}_i = [pa_{i_1}, pa_{i_2}, \cdots, pa_{i_K}]^T$, $i = 1, 2, \cdots, n$, is a K-dimensionality vector computed from Eq. (4) and n is the number of probe patches. $\mathbf{pa}$ is a $K \times n$ matrix.

We consider that the more ambiguous it is between $\mathbf{y}_i$ and $\mathbf{d}_j$ (i.e. larger pa_{i_j}), the less likely should $\mathbf{d}_j$ be selected for reconstructing $\mathbf{y}_i$. Therefore, we wish to find a coding vector $\mathbf{x}_i$ such that $\mathbf{x}_i$ is positively correlated to $-\mathbf{pa}_i$. This is formulated by the following sparse coding model:

$$\min_{\mathbf{x}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 + \alpha \mathbf{pa}_i^T \mathbf{x}_i, i = 1, \cdots, n, \tag{6}$$

where $\alpha \geq 0$. In practice, one can further constrain the sparsity of coding vector $\mathbf{x}_i$, since few gallery patches should be used for reconstructing each probe patch. Our sparse representation formulation thus becomes:

$$\min_{\mathbf{x}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{x}_i\|_2^2 + \alpha \mathbf{pa}_i^T \mathbf{x}_i + \beta \|\mathbf{x}_i\|_1, \; i = 1, \cdots, n, \tag{7}$$

where $\beta \geq 0$ constrains the strength of the sparsity constraint. Let $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_n]$, columns of which are the features of the n patches contained in a probe image and $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n] \in \mathbb{R}^{K \times n}$ is the corresponding ambiguity coefficient (coding) matrix. Similar to sparse representation classification, we classify a probe partial person image $\mathbf{Y}$ to class $\hat{c}$ by

$$\hat{c} = \arg\min_c r_c(\mathbf{Y}) = \frac{1}{n} \sum_{i=1}^{n} \|\mathbf{y}_i - \mathbf{D}_c \delta_c(\mathbf{x}_i)\|_2^2 \,, \tag{8}$$

where δ_c is a function that selects the coefficients associated with the c^{th} class. The sum of reconstruction residuals of n patch features is the basis on which we determine the identity. We call the above classifier as an ambiguity-sensitive matching classifier (AMC).

Discussions. Our local-to-local model is designed to cope with severe occlusions, and is intrinsically insensitive to mis-alignment and the presence of background patches. In particular, the proposed ambiguity-sensitive matching is formulated precisely to address the individual local patch mis-alignment problem and minimise the distraction caused by the background patches. This is because it essentially does a patch-to-set-of-patches matching, optimised against the reconstruction error, thus avoiding the error-prone

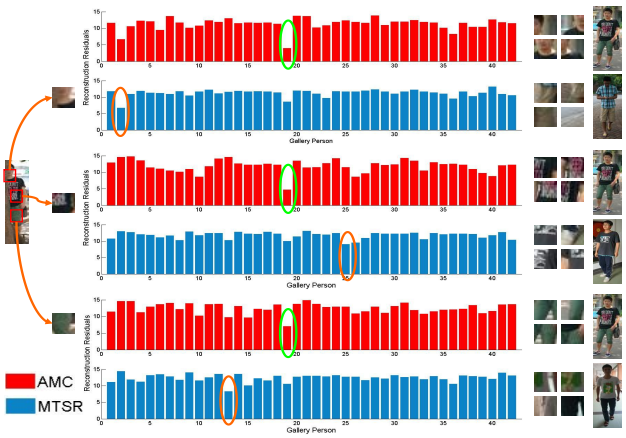


Figure 4. An example of our AMC model used for partial person matching. From left to right are the partial observation annotated by operators, the reconstruction error based on gallery patches of each person, the retrieved more similar gallery patches, and the matched gallery person image. It shows that our proposed method (AMC, denoted in red bar) can retrieve more similar patches from the right gallery person image against clutter background, compared to the standard sparse method without ambiguity modelling (MTSR [17], denoted in blue) .

patch-to-patch matching. Background patches are automatically excluded due to their low visual similarity and high ambiguity score. In Fig. 4, an example of local-to-local matching using our model is shown to demonstrate how our approach can effectively handle severe occlusions and the effects of background patches.

Optimisation of Ambiguity-Sensitive Matching. Minimising Eq. (7) is equivalent to minimising the following

$$\min_{\mathbf{x}_i} \frac{1}{2}\mathbf{x}_i^T\mathbf{D}^T\mathbf{D}\mathbf{x}_i - (\mathbf{D}^T\mathbf{y}_i - \frac{\alpha}{2}\mathbf{p}\mathbf{a}_i)^T\mathbf{x}_i + \frac{\beta}{2}\|\mathbf{x}_i\|_1. \quad (9)$$

Let $\mathbf{A} = \mathbf{D}^T\mathbf{D}$, $\mathbf{B} = -(\mathbf{D}^T\mathbf{y}_i - \frac{\alpha}{2}\mathbf{p}\mathbf{a}_i)$, $C = \frac{\beta}{2}$. Then, Eq. (9) is rewritten as

$$\hat{\mathbf{x}_i} = arg\min_{\mathbf{x}_i} \frac{1}{2}\mathbf{x}_i^T\mathbf{A}\mathbf{x}_i + \mathbf{B}^T\mathbf{x}_i + C\|\mathbf{x}_i\|_1. \quad (10)$$

In this work, we utilise the feature-sign search algorithm [11] to solve the optimisation problem in Eq. (10). The algorithm is summarised in Algorithm 1.

3.2. Global-to-Local Matching

The local patch-based matching model presented above does not capture the spatial layout information of local patches. In particular, there is no guarantee that given a partial person probe image, the matched gallery local patches form a coherent and compact regions corresponding to the same body parts in the probe. To overcome this limitation, we further consider a sliding window matching (SWM) process to perform global-to-local matching, i.e. the matching between the whole partially observed appearance of a probe person image and any local portion of a gallery image. Specifically, given a probe partial image of a person, we represent it using the same features as used in the last section. We set up a sliding window of the same size as the probe image. We then search for the most similar image region within each gallery image by sliding the window in each of the gallery images (the sliding search step is 5 pixels in our experiments). We use L_1-norm to measure the distance between the probe and the gallery image region within the sliding window. We compute a minimum distance l_c for the c-th class of gallery images. Therefore a minimum distance vector for all the C classes is denoted as

$$\mathbf{L}_{dist} = [l_1, l_2, \cdots, l_C]^T. \quad (11)$$

Subsequently, an unknown probe image can be identified as class $\hat{c} = \arg\min_c l_c,\ c = 1, 2, \cdots, C$.

Algorithm1 : The Optimisation Algorithm for AMC

Initialise: $\mathbf{x}_i = \mathbf{0}$, $\mathbf{s} = \mathbf{0}$ where $s_j \in \{-1, 0, 1\}$ denotes sign(x_{ij}), and active set $\Omega = \{\}$.
1: **while** 1 **do**
2: Compute $\mathbf{G} = \frac{\partial(\frac{1}{2}\mathbf{x}_i^T\mathbf{A}\mathbf{x}_i + \mathbf{B}^T\mathbf{x}_i)}{\partial \mathbf{x}_i} = \mathbf{A}\mathbf{x}_i + \mathbf{B}$.
3: From zero coefficients of $\mathbf{x}_i$, select $j = arg\max_j |G_j|$.
4: If $G_j > C$, then set $s_j = -1$, $\Omega = \Omega \cup \{j\}$.
5: If $G_j < -C$, then set $s_j = 1$, $\Omega = \Omega \cup \{j\}$.
6: **while** 1 **do**
7: Let $\hat{\mathbf{A}}$ contain intersections between columns and rows of $\mathbf{A}$ corresponding to Ω. Let $\hat{\mathbf{B}}$, $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{s}}$ be a subvectors of $\mathbf{B}$, $\mathbf{x}_i$ and $\mathbf{s}$ corresponding to Ω.
8: For the Quadratic Programming($\min_{\hat{\mathbf{x}}_i} \frac{1}{2}\hat{\mathbf{x}}_i^T\hat{\mathbf{A}}\hat{\mathbf{x}}_i + \hat{\mathbf{B}}^T\hat{\mathbf{x}}_i + C\hat{\mathbf{s}}^T\hat{\mathbf{x}}_i$), compute the analytical solution $\hat{\mathbf{x}}_i^{new} = -\hat{\mathbf{A}}^{-1}(\hat{\mathbf{B}} + C\hat{\mathbf{s}})$.
9: Perform a discrete line search on the closed line segment from $\hat{\mathbf{x}}_i$ to $\hat{\mathbf{x}}_i^{new}$: Check the objective value at $\hat{\mathbf{x}}_i^{new}$ and all points where any coefficient changes sign, then update $\hat{\mathbf{x}}_i$ and the corresponding entries in $\mathbf{x}_i$ to the point with the lowest objective value.
10: Remove 0 coefficients of $\hat{\mathbf{x}}_i$ from Ω and update $\mathbf{s}$.
11: If $G_l + Cs_l = 0, l \in \{l|x_{il} \neq 0\}$, break.
12: **end while**
13: If $|G_l| \leq C, l \in \{l|x_{il} = 0\}$, break and return $\mathbf{x}_i$ as the solution.
14: **end while**

3.3. Classification

Given a probe partial person image with n patches, a sum fusion among reconstruction residuals of all patch features with respect to each class is calculated from Eq. (8). It can be denoted as $\mathbf{R}_{dist} = [r_1, r_2, \cdots, r_C]^T$. We combine it with the distance vector in Eq. (11). The final distance vector can be written as

$$\mathbf{S}_{dist} = \gamma\mathbf{R}_{dist} + (1 - \gamma)\mathbf{L}_{dist}, \quad (12)$$

where γ is a weight for regulating the effect of local distance and global distance. Finally, the identity of the probe image can be determined by $\hat{c} = \arg\min_c s_c$, where s_c is the c^{th} entry of $\mathbf{S}_{dist}$. We denote the above fusion model for combining our ambiguity-sensitive matching classifier (AMC) and sliding window matching (SWM) as AMC-SWM.

4. Experiments

4.1. Datasets and Settings

New Partial REID Dataset. There is no partial person re-id dataset publicably available. To fill this gap, we contribute a new partial person dataset named Partial REID. The dataset includes 600 images of 60 people, with 5 full-body images and 5 partial images per person. The images were collected at an university campus with different viewpoints, background and different types of severe occlusions (see Fig. 2). The new dataset will be released on our website soon.

Two Simulated Datasets. Two simulated partial person datasets named P-iLIDS and P-CAVIAR were also used for evaluation, which are based on i-LIDS [40] and CAVIAR [3], respectively. In the i-LIDS dataset, there are 119 people with total 476 person images captured by multiple non-overlapping cameras. Different from most other datasets, the original images have fair amount of occlusion, sometimes rather severe, caused by people and luggage. For each person, we generated its partial observation by selecting the most occluded image of that person and then manually crop the un-occluded part of the same person image to create the P-iLIDS dataset (see Fig. 5). The CAVIAR dataset contains 1220 images of 72 individuals from 2 cameras in a shopping mall. We randomly selected half of holistic images of each person to generate the partial images (see Fig. 5).

Compared Methods. For comparison, six existing representative re-id models were considered, including the relative distance comparison (RDC) method [42], the ranking based model using PRSVM [26], the local fisher discriminant analysis (LFDA) [25], the KISSME distance learning method [9], locally-adaptive decision functions (LADF) [15] and the non-learning distance based L1-norm matching. However, existing re-id methods are not designed for solving the partial person re-id; they are thus not expected to be competitive. Hence, we also selected some representative models that address the related partial face recognition problem. These include the multi-task sparse representation (MTSR) [17] proposed for partial face modelling, and the standard sparse representation classifier (SRC) [32] which is a general-purpose classifier but has obtained strong performance on recognising face images with occlusion.

Features. All methods were evaluated using the same set of features for fair comparison. Specifically, features were extracted from a 16×16 support area, and these support areas were densely sampled with an overlap of half of the width/height of the supporting area in both horizontal and vertical directions. Each region was represented by a 16-bin histogram of 8 colour channels (RGB, YCbCr, HS), uniform LBP histograms and HOG descriptors. So each patch was represented by a 484-dimensional feature vector. The partial image size was normalised to 128×48 pixels for the compared re-id methods in our experiments, and a total of 75 regions were selected, resulting in a 36300-dimensional vector. In contrast, the image size was not normalised in our method. Each image was represented by a feature matrix, and each column is a feature vector of each 16×16 patch, resulting in a $484 \times m$ matrix, where m is the number of regions per image and may be different for different images.

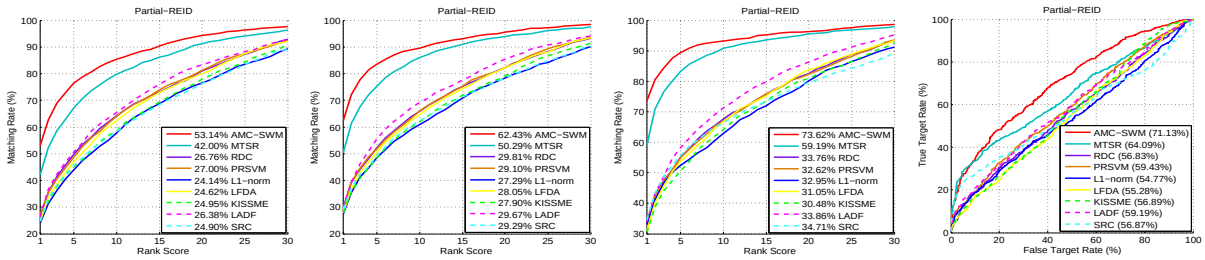
Figure 5. Examples of partial person images (first row) and the corresponding full images (second row). From left to right, columns 1–3 are from P-iLIDS, and columns 4–6 from P-CAVIAR.

Settings. Both single-shot and multi-shot experiments were conducted. The test sets of all datasets were randomly selected using 70% of the individuals. Specifically, there are p = 42, p = 50, and p = 83 individuals in each of the test sets for the Partial REID, P-CAVIAR and P-iLIDS datasets, respectively. Each test set was composed of a probe set and a gallery set. The probe set consists of all partial images per person, and the holistic person images were used as the gallery set. This procedure was repeated 10 times. For evaluation, we used the average cumulative match characteristic (CMC) curves to measure the matching performance.

In most conventional person re-id works and most experiments in this work, the gallery vs. probe is in a closed-set setting, i.e. the probe and gallery contain exactly the same set of people. To see how our algorithm performs against imposters, we also provide the ROC curves for open-set testing, under which images of 30% of the gallery people were randomly removed in each closed-set testing case. Due to the space limitation, we only provided the ROC curves in the single-shot experiments.

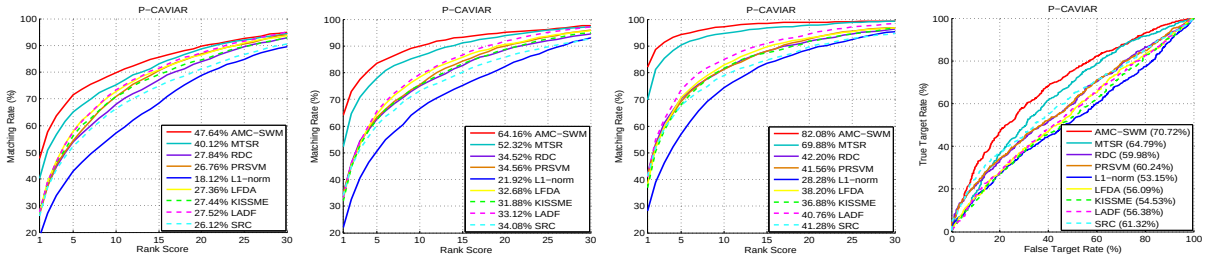
4.2. Evaluations on the Partial REID Dataset

Single-shot Experiments. Single-shot re-id means that a single (N=1) image is used as the gallery image for each person. Firstly, we compared our AMC-SWM against existing re-id methods including PRSVM, RDC, L1-norm, LFDA, KISSME and LADF. The CMC curves in Fig. 6 show clearly that much worse performance is obtained when using these conventional re-id methods compared to our method. This is because that the same person's appearance will experience dramatic changes in the partial re-id setting and these methods assume full body detection for both the probe and gallery images. They thus all perform more poorly on this more challenging partial re-id task. We then compared the proposed AMC-SWM with the two representative partial face recognition methods MTSR and SRC. Among the two, MTSR is closely related to our AMC model because



(a) N=1,CMC (b) N=2,CMC (c) N=5,CMC (d) N=1,ROC

Figure 6. Performance comparison on the PARTIAL REID dataset: CMC curves with rank-1 matching rate, and ROC curves with area-under-curve (AUC) values (p = 42, best viewed in colour)

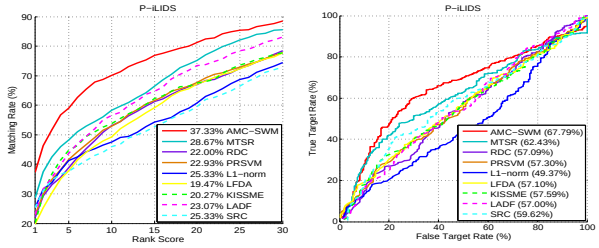


(a) N=1,CMC (b) N=2,CMC (c) N=5,CMC (d) N=1,ROC

Figure 7. Performance comparison on P-CAVIAR dataset: CMC curves and ROC curves (p = 50)

it also employs sparse coding for local patch modelling. The main difference is that AMC incorporates the ambiguity score to guide the local-to-local patch-based matching. Figure 6 shows that MTSR is much more competitive than the six compared full-body person re-id models. However, our AMC still achieves at least 3% higher matching rate at rank-1 and our whole algorithm AMC-SWM outperforms MTSR by a large margin by comparing Table 1 and Fig. 6. In contrast, the SRC model yields very weak performance. This is because although SRC is also a sparse-based classifier, it directly uses the entire input partial images rather than local patches. It is thus sensitive to the severe misalignment problem. Similar results are also observed on the ROC curves under the open-set setting (see Fig. 6(d)).

Multi-shot Experiments. We also evaluated our AMC-SWM method under the multi-shot setting (N=2 and N=5) on the Partial REID dataset. The results are shown in Fig. 6. Overall, similar results are obtained as in the single-shot experiments. In particular, the results show that multi-shot helps, since the matching rate of AMC-SWM increases from 53.14% (N=1) to 73.62% (N=5) at rank-1. In contrast, the performance improvement to the other existing re-id methods are also more modest when compared to that of the AMC-SWM model. For instance, KISSME increases from 24.95% (N=1) to 30.48% (N=5) at rank-1. Although MTSR also performs better when more samples are used, the gap between MTSR and AMC-SWM increases at rank-1 score (a 11% difference at N=1 was increased to 14% at N=5). Again the SRC which performs directly on input partial images does not perform well.



(a) N=1,CMC (b) N=1,ROC

Figure 8. Performance comparison using CMC and ROC curves on the P-iLIDS dataset (N=1), (p = 83)

4.3. Evaluations on the Simulated Datasets

The proposed AMC-SWM was also compared with the existing methods on the two simulated partial re-id datasets P-CAVIAR and P-iLIDS. The results are shown in Figs. 7 and 8. Since there is always only one pair of images for each person in i-LIDS, we only report single-shot results (N = 1). It is evident that all the observations from the results on the Partial REID dataset experiments remain valid for these two simulated datasets. For instance, on P-CAVIAR, AMC-SWM gains 7.5% at N = 1 and 12% at $N >= 2$ for rank-1 performance over the second best performed model MTSR. The gap become even bigger on P-iLIDS which contains more naturally occluded person images than P-CAVIAR.

4.4. Further Analysis

Contributions of Individual Components. Our proposed method consists of two matching components: AMC and

Methods	N=1			N=2			N=5			
	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	
AMC-SWM	**53.14**	**76.43**	**85.29**	**62.43**	**83.62**	**89.57**	**73.62**	**89.43**	**93.29**	Partial REID
AMC	45.19	70.29	81.00	53.48	77.43	87.19	63.86	86.57	91.48	$p=42$
SWM	47.24	71.05	80.57	56.00	77.67	86.10	65.24	85.10	90.91	
KLFDA	23.47	47.43	60.90	27.00	49.76	63.05	30.90	52.95	65.67	
AMC-SWM	**47.64**	**71.52**	**79.80**	**64.12**	**83.44**	**89.92**	**82.08**	**94.40**	**97.24**	P-CAVIAR
AMC	44.72	67.52	78.32	59.48	82.00	88.68	78.84	92.20	95.56	$p=50$
SWM	44.16	65.08	74.72	58.88	78.12	85.96	76.92	91.28	95.32	
KLFDA	27.88	57.00	70.92	33.28	64.04	77.72	37.96	71.52	84.36	
AMC-SWM	**37.33**	**59.07**	**70.13**	-	-	-	-	-	-	P-iLIDS
AMC	31.87	52.80	59.60	-	-	-	-	-	-	$p=83$
SWM	35.73	54.53	65.87	-	-	-	-	-	-	
KLFDA	20.27	37.33	47.73	-	-	-	-	-	-	

Table 1. Further Analysis on AMC-SWM

SWM. In Table 1, we evaluated the two components on how they contribute to the full model. The results show that both of them are effective on their own (each outperforms all the compared methods). Moreover, when combined, the best performance is achieved. This validates our design consideration in that the two matching components are complementary and should be combined.

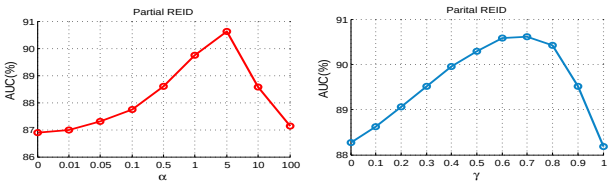
Evaluation on Effect of Kernel-based Method. Recently, Kernel LFDA (KLFDA) [33] achieved good performance for conventional person re-id [33] indicating that kernelisation helps. For partial person re-id problem, we additionally evaluated its effect in Table 1. Compared to LFDA in Figs. 6, 7 and 8, KLFDA does not differ from LFDA on the performance for partial re-id. This suggests that the effect of kernelisation diminishes for partial person matching.

Influence of Parameters. We evaluated two key parameters in our modelling, the strength of the ambiguity modelling (i.e. α in Eq. (7)) and the weight to combine two matching models (i.e. γ in Eq. (12)). Due to space limitation, we only show results on the Partial REID dataset, similar results were obtained from the other datasets. As shown in Fig. 9, overall the effects of both parameters are small. When α is approximately 5 and γ is around 0.7, the proposed AMC-SWM model achieves the best performance.

4.5. Discussions

Our experiments show that the existing person re-id methods including RDC, KISSME and LADF perform poorly for partial person re-id when compared to the proposed AMC-SWM model. This is expected because these methods assume that full body appearance of a person is available for matching. When this assumption is invalid, there is no mechanism in these models that can cope effectively with the challenges of matching a partial observation against a full observation across camera views.

A more sensible solution is to build a model specifically designed for matching the partial observation against part of a full observation. Such a model does not exist in existing person re-id works but has plenty of options in the related field of partial face recognition. Indeed, the MTSR model, originally designed for partial face recognition, has been



(a) Performances w.r.t.α. (b) Performances w.r.t.γ.

Figure 9. Evaluation of different parameter (Eqs. (7)& (12)) using AUC curves on PARTIAL REID (N=1).

shown in our experiments to be more effective than other full-body based re-id methods. However, without considering person re-id specific constraints, MTSR is notably inferior to the proposed AMC model. This suggests that there are far great ambiguity in the person body patch appearance. Particularly, many utility patches exist which can be combined to approximate any given patch in a probe image. The ambiguity score introduced in our AMC model is specifically designed to overcome this patch ambiguity problem by preventing visually dissimilar gallery patches from being used for reconstructing the probe patch. This contributes to the improvement of the proposed AMC over MTSR. Finally, the best performance of the proposed full AMC-SWM model suggests that combining the global and local matching models in this paper is the optimal solution as it exploits the complementarity of the two models.

5. Conclusion

We have considered a new and more realistic person re-identification challenge: the partial person re-identification problem. To address the unique challenges associated with this new re-id problem, we proposed a novel framework that consists of a local-to-local matching model and a global-to-local matching model. Both models are specifically designed to address the partial re-id problem and are complementary to each other. The effectiveness of our models has been demonstrated by extensive experiments on three datasets including a new partial person re-id dataset introduced in this paper. Future work includes extending our model to handle multi-scale images of body parts when they are are of very different scales. As a simple solution, the image of each scale can be matched using the proposed method, followed by fusing the matching scores or selecting the scale with the highest score.

Acknowledgment

This work was supported partially by the National NSFC (No. 61472456), NSFC for Excellent Young Scientist Programme (No. 61522115), Guangzhou Pearl River Science and Technology Rising Star Project (No. 2013J2200068), and in part by the Guangdong Natural Science Funds for Distinguished Young Scholar (No. S2013050014265).

References

[1] A. M. Andrés, S. Padovani, M. Tepper, and J. JacoboBerlles. Face recognition on partially occluded images using compressed sensing. *PRL*, 36, 2014.

[2] D. Baltieri, R. Vezzani, and R. Cucchiara. Sarc3d: A new 3d body model for people tracking and reidentification. In *ICIAP*. 2011.

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

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

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

[6] R. He, W.-S. Zheng, and B.-G. Hu. Maximum correntropy criterion for robust face recognition. *IEEE TPAMI*, 33(8), 2011.

[7] J. Hu, J. Lu, and Y.-P. Tan. Robust partial face recognition using instance-to-class distance. In *VCIP*, 2013.

[8] B. Huang, J. Chen, Y. Wang, C. Liang, Z. Wang, and K. Sun. Sparsity-based occlusion handling method for person reidentification. In *MultiMedia Modeling*. 2015.

[9] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *CVPR*, 2012.

[10] I. Kviatkovsky, A. Adam, and E. Rivlin. Color invariants for person reidentification. *IEEE TPAMI*, 35(7), 2013.

[11] H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In *NIPS*, 2006.

[12] Y. J. Lee and Y. T. Park. Occluded 3d object recognition using partial shape and octree model. In *Advances in Intelligent Computing*, 2005.

[13] W. Li, R. Zhao, and X. Wang. Human reidentification with transferred metric learning. In *ACCV*, 2013.

[14] W. Li, R. Zhao, T. Xiao, and X. Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *CVPR*, 2014.

[15] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In *CVPR*, 2013.

[16] G. Lian, J. Lai, and W.-S. Zheng. Spatial-temporal consistent labeling of tracked pedestrians across non-overlapping camera views. *PR*, 44(5), 2011.

[17] S. Liao, A. K. Jain, and S. Z. Li. Partial face recognition alignment-free approach. *IEEE TPAMI*, 35(5), 2013.

[18] G. Lisanti, I. Masi, A. Bagdanov, and A. Del Bimbo. Person re-identification by iterative re-weighted sparse ranking. *IEEE TPAMI*, 2014.

[19] C. Liu, C. C. Loy, S. Gong, and G. Wang. Pop: Person re-identification post-rank optimisation. In *ICCV*, 2013.

[20] B. Ma, Y. Su, and F. Jurie. Local descriptors encoded by fisher vectors for person re-identification. In *ECCV Workshop*, 2012.

[21] L. Ma, X. Yang, and D. Tao. Person re-identification over camera networks using multi-task distance metric learning. *IEEE TIP*, 2014.

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

[23] R. Min, A. Hadid, and J. Dugelay. Improving the recognition of faces occluded by facial accessories. In *FG*, 2011.

[24] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, and Z. Zhu. Robust face recognition via occlusion dictionary learning. *PR*, 47(4), 2014.

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

[26] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *BMVC*, 2010.

[27] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li. Person re-identification by regularized smoothing kiss metric learning. *IEEE TCSVT*, 23(10), 2013.

[28] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *CVPR*, 2011.

[29] T. Wang, S. Gong, X. Zhu, and S. Wang. Person re-identification by video ranking. In *ECCV*, 2014.

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

[31] R. Weng, J. Lu, J. Hu, G. Yang, and Y.-P. Tan. Robust feature set matching for partial face recognition. In *ICCV*, 2013.

[32] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE TPAMI*, 31(2), 2009.

[33] F. Xiong, M. Gou, O. Camps, and M. Sznaier. Person re-identification using kernel-based metric learning methods. In *ECCV*, 2014.

[34] Y. Xu, L. Liang, W.-S. Zheng, and X. Liu. Human re-identification by matching compositional template with cluster sampling. In *ICCV*, 2013.

[35] M. Yang and L. Zhang. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In *ECCV*, 2010.

[36] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *ICCV*, 2013.

[37] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *CVPR*, 2013.

[38] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person re-identification. In *CVPR*, 2014.

[39] W.-S. Zheng, S. Gong, and T. Xiang. Associating groups of people. In *BMVC*, 2009.

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

[41] W.-S. Zheng, S. Gong, and T. Xiang. Transfer re-identification: From person to set-based verification. In *CVPR*, 2012.

[42] W.-S. Zheng, S. Gong, and T. Xiang. Re-identification by relative distance comparison. *IEEE TPAMI*, 35(3), 2013.

[43] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma. Face recognition with contiguous occlusion using markov random fields. In *CVPR*, 2009.