IEEJ TRANSACTIONS ON ELECTRICAL AND ELECTRONIC ENGINEERING
IEEJ Trans 2020; **15**: 909–917
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI:10.1002/tee.23133

Paper

Simultaneous Image Matching for Person Re-identification via the Stable Marriage Algorithm

Nik Mohd Zarifie Hashim[*,**a], Non-member
Yasutomo Kawanishi[*], Non-member
Daisuke Deguchi[***], Non-member
Ichiro Ide[*], Non-member
Hiroshi Murase[*], Non-member

Recently, multiperson tracking across non-overlapping camera views is gaining increasing attention in the video surveillance field due to its importance for public security. Most existing methods find a correspondence of person images individually across each pair of adjacent two cameras. However, in such a naive approach, redundant matching where one of the person images is selected for the matching pair for several times, often occurs. It leads to tracking failure. To overcome this problem, in this paper, by considering the matching problem as an instance of the "marriage problem" that is well known in Economics, we propose a simultaneous image matching method for multiple persons based on the stable marriage algorithm. We confirmed that the proposed method outperforms some of the existing state-of-the-art methods on several well-known public datasets.

Keywords: person re-identification; simultaneous matching; stable marriage algorithm

Received 20 August 2019; Revised 9 February 2020

1. Introduction

Video surveillance covering a broad area is crucial for protecting the public security. The increment of crime in public areas accelerates the usage of camera surveillance for capturing videos especially after the 911 terrorists' attack [1]. For tracing a targeted person who commits a crime, finding the trajectory of the person across multiple cameras is essential. Obtaining trajectories of customers across multiple camera views in a shopping mall is also an important task, namely, where they visited and which route they selected is valuable for marketing.

Trajectories of all of the people in surveillance videos should be estimated by tracking. Tracking people across camera views can be done by matching person images captured at adjacent camera views. The problem is called person re-identification. For this, there are several issues including occlusion and illumination variations. Thus, a robust matching scheme is a crucial element for matching them.

Typically, in the person re-identification problem setting, a set of person images captured by one camera-view is considered as a query image set and that captured from another camera-view is considered as a gallery image set. There is various related work that employs various image features and distance metrics for solving the person re-identification problem. However, there is still a lack of work which focuses on the matching scheme for person re-identification. In general, there are two types of image matching schemes for person re-identification, individual and simultaneous image matching.

The individual matching scheme finds the matching image pair independently based on the similarity between the query to the gallery image. The independent image matching allows the previous mismatched gallery image with a query image to be matched again with another query image. The majority of the conventional person re-identification methods employ the former scheme, individual image matching. Since the justification of the matched image pair is based only on image similarity, mismatch of images occurs [2]. Once a mismatch occurs, redundant matchings where one gallery image is matched with multiple query images will occur subsequently. In the individual image matching scenario, when a redundant image matching occurs, we need to select one of the matched images as the tracking result. In this case, the matching results other than the selected one are discarded. These discarded images will remain as unmatched images that will be matched with other unmatched images. Because of this, redundant matching leads to tracking failure.

Figure 1 illustrates an example of the former image matching scheme for a query image set and a gallery image set. Each query image is compared with the gallery images by the image features and all the gallery images are ranked by the distance. Here, a_1 and a_2 from the query image set are matched to the same b_3 in the gallery image set. Only one match from these two is selected as a tracking result, e.g., the pair a_1–b_3. The remaining matched pair a_2–b_3 is discarded. Meanwhile a_3 is matched to b_1. As a result of the individual matching, only a_3 is successfully matched by the matching of the pair a_3–b_1, while a_1 is unsuccessfully matched by the matching of the pair a_1–b_3, and a_2 is not matched at all. The redundant matching influences the whole image matching results with unmatched images.

Meanwhile, the latter image matching scheme finds the corresponding images for all query images simultaneously. This

[a] Correspondence to: Nik Mohd Zarifie Hashim. E-mail: hashimz@murase.is.i.nagoya-u.ac.jp

[*]Graduate School of Informatics, Nagoya University, Nagoya, Japan
[**]Centre for Telecommunication Research & Innovation (CeTRI), Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia
[***]Information Strategy Office, Nagoya University, Nagoya, Japan

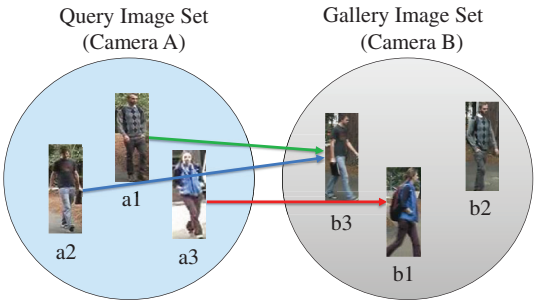

Fig. 1. Example of mismatch by individual image matching on the VIPeR dataset [3]

simultaneous image matching is similar to the concept of two sets of element matching in a graph [4]. Here, the concept of two sets element matching is similar with the two image sets matching; query and gallery images, in person re-identification which was introduced in Ref. [2] Since this can be considered as a bipartite graph matching problem, we consider the simultaneous image matching scheme as the bipartite graph matching problem to solve the person re-identification problem.

There are several algorithms for bipartite graph matching; Greedy matching (GM), and Hungarian matching (HM) [5]. GM simply matches by finding the maximum weight of the images one-by-one [6,7]. The GM helps us to find the maximum value of image similarity. The indirect involvement of GM for person re-identification was proposed as a pre-processing step in their tracking algorithm [8] and as a patch matching tool for solving their feature-shift problem [9]. In person re-identification, this simple concept of matching the images without any replacement allows the existence of an unmatched image; the matched image will not be rematched again. Although it is popular as an optimizing tool for various mathematical problems [10,11], it is still not deployed in existing person re-identification methods as the majority of them still implement the traditional image matching, individual image matching.

Recently, Zhang *et al.* proposed a method named person re-identification via structured matching (PRiSM). They deployed the weighted structured matching method [2] to the person re-identification problem by considering it as an instance of structured matching. A number of re-identification research studies assume the number of images in both image sets are the same. This assumption makes the re-identification problem simple. Zhang *et al.* also tackle a more complicated problem when the numbers of images in both cameras are not the same and even if the number is equal, some persons in a camera view are replaced by other persons.

In this paper, we propose a new image matching scheme for person re-identification by considering the image matching problem as the marriage problem well known in Economics [12]. As the marriage problem lists elements in the other set in terms of the preference, the proposed method lists the person images with a ranking order of similarities.

Besides considering the problem of person re-identification in the same number of images in both cameras, in this paper, we extended the idea of [2] with several person re-identification datasets. This study is essential to show the robustness of the proposed method in various cases of different numbers of images.

The main contributions of this research can be summarized as follows:

- We introduce a simultaneous person matching scheme for person re-identification based on the stable marriage algorithm (SMA) that is well known in Economics.
- We provide an analysis on a comparison of the image matching methods in cases where the image sets contain a different numbers of images.

Note that this paper is an extended version of our previous conference paper [13]. We extended the paper by adding more detailed explanation of the proposed method and further evaluation on various publicly available datasets.

The remaining part of this paper is structured as follows: In Section 2, we will discuss the related work. In Section 3, we will explain the proposed method which uses SMA in detail. Section 4 will discuss the evaluation and the results. In the end, we will conclude the paper in Section 5.

2. Related Work

In the past few years, many researchers proposed various methods to improve the person re-identification performance. Illumination variation, human pose variation, and occlusion are the common difficulties in conventional person re-identification research studies. In this section, we briefly introduce some of the representative methods which are related to our work. In this paper, we divide existing methods into the following two groups: individual image matching and simultaneous image matching.

2.1. Individual image matching

The majority of the existing person re-identification methods focus on matching images individually. The idea of the individual image matching scheme is that each query image is matched with an image in the gallery set individually. This scheme matches all the images in query images with gallery images without any replacement of the other matching results. Matching without any replacement leads to redundant matchings where a gallery image is matched several times to the query images.

The majority of the previous work utilizes the variety of image features for image matching [14–29]. Farenzena *et al.* proposed a feature aggregation method that focused on the symmetry of a human body. Recent methods in person re-identification tend to focus on metric learning. By utilizing metric learning, better similarity metrics are obtained [30–38]. Ahmed *et al.* initiated person re-identification with Deep Learning by proposing a CNN-based method [39]. Liu *et al.* proposed a triplet-loss-based CNN model to find useful features and metrics [40]. A similar approach also has been introduced in Ref. [41] Applying Deep Learning in person re-identification, however, struggles with the image labeling problem in the training process. It will consume more time for manual labeling of training data.

2.2. Simultaneous image matching

This kind of person re-identification scheme matches all given pedestrian images simultaneously, considering all matching results. The simultaneous image matching scheme is usually formulated as a bipartite graph matching problem. A bipartite graph consists of given two sets of vertices and edges, which connect the two vertices. The bipartite graph matching problem finds a set of edges of a bipartite graph which maximizes the total scores corresponding to the edges. For person re-identification, the score of an edge can be considered as a similarity score between two images, when a person is considered as a vertex in V of a graph $G = (V, E)$.

The most naive approach to find an optimal solution for this is GM. Its algorithm starts from giving scores of connecting two vertices in different sets, and then adds edges to the vertex pair whose vertices are not connected to any vertex yet in descending order. Another well-known solution is HM [5]. Despite being the choice in a particular field, many works consider the HM as a tool for optimization at the back end. [11,42]

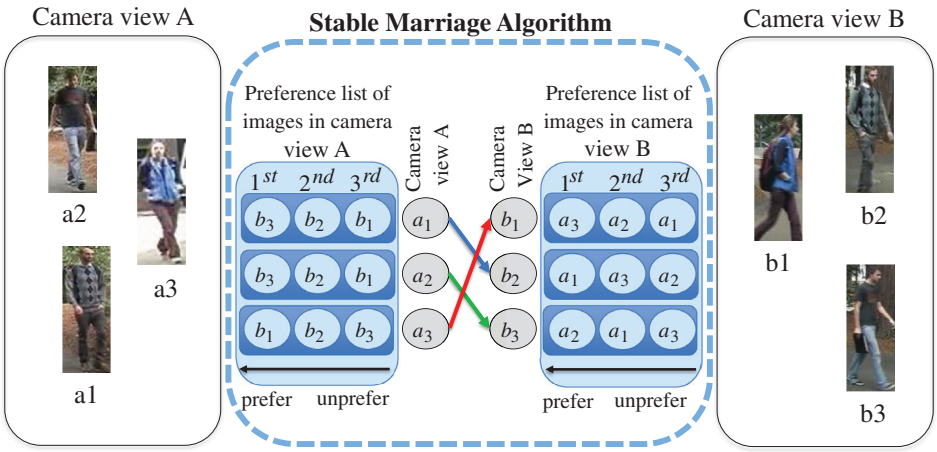

Fig. 2. Overview of the proposed method. We consider the person image as an instance of the stable marriage problem. The gray circles represent both camera views' person images as an element of each set, while the light blue circles in the blue shaded areas represent preference lists sorted horizontally for each camera view from preferable to unpreferable. The stable marriage algorithm matches images from camera view A's image set to camera view B's image set based on the preference list. The blue line represents the matching from person a_1 to b_2, the green line from a_2 to b_3, and red line from a_3 to b_1

Ye *et al.* proposed another idea in simultaneous image matching. They initiated the idea of structured learning in graph matching for person re-identification [43]. This idea was extended and improved by Zhang *et al.* [2] named PRiSM.

3. Person Re-identification via the SMA

3.1. Overview Here, we assume that every person is detected by a camera-view without redundancy. By considering the person re-identification problem as an instance of the marriage problem, we propose a matching method which finds a stable matching between two image sets. To solve the image matching in person re-identification, we introduce SMA for person re-identification. Most of the conventional person re-identification methods consider only similarities of images but do not consider the similarity ranking from both sides. Meanwhile, the proposed method considers the similarity ranking, which allows robust matching.

Thanks to the algorithm, we realize a bipartite graph matching-based person re-identification method. Overview of the proposed method is illustrated in Fig. 2. Since the matching part is the core issue in this paper, we first explain the matching process. After that, we will explain the details of feature extraction and similarity distance calculation in the following subsections. In this paper, selecting image features is not our main focus, so we just use simple image features, while there are more sophisticated features proposed in state-of-the-art person re-identification methods.

3.2. Marriage problems The marriage problem was originally proposed by Gale and Shapley [12] a few decades ago in Economics. The concept, which described a matching solution for elements in two sets of elements given a ranking of preference for each element, was successfully implemented in a real-life situation; doctor-hospital assignment in the United States of America in 1962 [44]. The marriage problem is composed by the following two definitions. In the definitions, we assume that there are men 1, 2, . . . , and women 1, 2, . . . , each with a preference list of the opposite sex to propose to.

Definition 1. An unstable matching of marriage is defined as two persons; man 1 and man 2, are assigned to woman 1 and woman 2, respectively, although man 2 prefers woman 1 to woman 2 and woman 1 prefers man 2 to man 1.

Definition 2. The marriage problem is defined as *stable* if every person is matched with a partner.

SMA gives stable matching pairs of the instances by rematching an existing pair if it showed less rank preference by their partner. Thus, if there are the same numbers of men and women, each of them will have a partner based on his/her preference list. This stable marriage problem is known to have a polynomial solution and it also is one example of bipartite graph matching.

3.3. Simultaneous image matching via the SMA For person re-identification, the similarity rank of a query image to all its gallery images could be considered a similar state with an element with their preference list in the SMA. The similarity image rank represents quantitatively how close a query image to other gallery images is. For this purpose, we propose the idea of simultaneous image matching by considering the similarity image rank as the preference list in SMA.

We propose the simultaneous image matching via SMA by initiating the idea of similarity image rank for the person re-identification as its preference list. The idea of having a similarity rank of the images before the matching is illustrated in Algorithm 1. To the best of our knowledge, SMA uses two preference ranking lists for Camera View A to B and Camera View B to A. In other words, the matching with SMA for two camera views is considered from both camera view's preference rankings. However, the traditional individual image matching considers only the ranking for Camera View A to B with several criteria.

By utilizing the proposed simultaneous image matching, preference rankings are considered for both Camera A to B and Camera B to A. Compared to traditional GM and HM, the proposed method will not only sort the preference ranking and choose the best image in the rank as the final matched image but also the confirmation with the opposite preference ranking which is a requirement for SMA, which finally yields robust image matching.

3.4. Similarity ranking In this paper, we focus on two kinds of image features; one is the traditional hue, saturation, value (HSV) color histogram, and the other is the modified image feature, named mask-improved symmetry-driven accumulation of local features (MSDALF) to calculate the similarity between images in both query and gallery images. For the HSV color histogram, the histogram intersection method as same as that in Refs. [45,46] is

Algorithm 1 Simultaneous Image Matching via the Stable Marriage Algorithm for Person Re-identification

Inputs:
$\mathcal{I}^A \leftarrow$ Query for N Images , $\mathcal{I}^B \leftarrow$ Gallery for N Images

Phase 1 - Preference List as Similarity Ranking

for $i \leftarrow 1$ to N **do**
  Rank(I_i^A) $\leftarrow$ sort($[I_1^B, I_2^B, ..]$) by s(I_i^A, I_j^B) in descending order
  Rank(I_i^B) $\leftarrow$ sort($[I_1^A, I_2^A, ..]$) by s(I_i^B, I_j^A) in descending order
end for

Phase 2 - Simultaneous Image Matching

Initialize:
  $\forall I_i^A \in \mathcal{I}^A$ and $\forall I_i^B \in \mathcal{I}^B$ as *free*
while $\exists$ free I_i^A which still has an I_i^B to be matched with **do**
  $I_i^B \leftarrow$ first rank in I_i^A's list and I_i^A has not yet been matched
  if I_i^B is *free* **then**:
    (I_i^A,I_i^B) becomes matched
  else:
    $\exists$ pair (I_k^A,I_i^B) already matched
    if I_i^B is more similar to I_i^A than I_k^A **then**:
      I_k^A becomes free
      (I_i^A,I_i^B) becomes matched
    else:
      (I_k^A,I_i^B) is kept matched
    end if
  end if
end while

used for image comparison to calculate the distance between the images accordingly. Since the HSV color histogram is commonly used in extracting a feature from an image, we considered it as our baseline image feature.

Meanwhile, MSDALF is an extension of SDALF, which is a well-known image feature for person re-identification. While the traditional SDALF utilizes person image mask generated by Stel component analysis [47] to calculate image features, the MSDALF estimates the person image mask by using Mask R-CNN [48]. These masked images of the viewpoint invariant pedestrian recognition (VIPeR) dataset are illustrated in Fig. 3. This feature depends on the quality of the image mask; therefore, we named it MSDALF. Thanks to the finer person image masks, the MSDALF can extract better image features which capture the characteristics of a person.

An M-bin color HSV histogram or MSDALF image feature is calculated from two image as I_i^A and I_j^B for the similarity calculation, respectively, where

$$\mathbf{x}_i^A = f(I_i^A), \tag{1}$$

$$\mathbf{x}_j^B = f(I_j^B). \tag{2}$$

The image feature similarity of two images will be calculated by the similarity function $s(I_i^A, I_j^B)$ which is defined as

$$s(I_i^A, I_j^B) = s_f(\mathbf{x}_i^A, \mathbf{x}_j^B), \tag{3}$$

where $s_f(\mathbf{x}_i^A, \mathbf{x}_j^B)$ is the histogram-intersection defined as

$$s_f(\mathbf{x}_i^A, \mathbf{x}_j^B) = \sum_{k=1}^{M} \min(\mathbf{x}_{ik}^A, \mathbf{x}_{jk}^B). \tag{4}$$

Using this appearance similarity, we sort the images and then apply SMA to find the matched image pairs by simultaneous image matching.

3.5. Example of simultaneous image matching

We formulate person re-identification across two camera views as an instance of marriage problem given person images $\mathcal{I}^A$ and $\mathcal{I}^B$ detected from two camera-views (Cameras A and B), respectively, as follows.

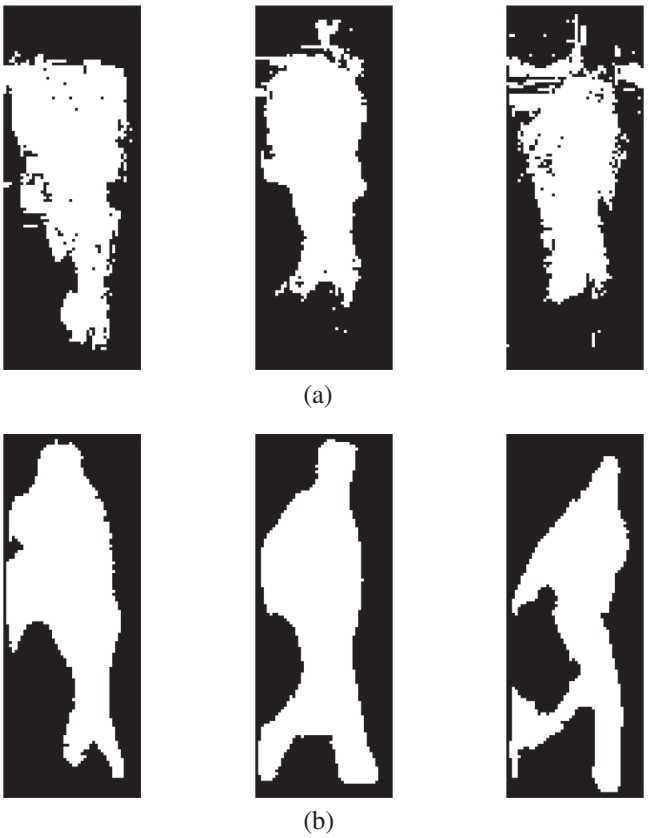

Fig. 3. Examples of mask images using the VIPeR dataset [3]

$$\mathcal{I}^A = \{I_1^A, I_2^A, \ldots \ldots, I_N^A\} \tag{5}$$

$$\mathcal{I}^B = \{I_1^B, I_2^B, \ldots \ldots, I_N^B\} \tag{6}$$

By following the SMA, we obtain a similarity ranking list for each person image. Table I shows three images from Camera A and three images from Camera B. For matching these images, for each image in $\mathcal{I}^A$ and images $\mathcal{I}^B$ are sorted by the similarity with respect to each image in $\mathcal{I}^A$ in descending order. The proposed algorithm, simultaneous image matching via SMA, is shown in Algorithm 1.

Table I. Example of image similarity based on feature similarity

(a) From $\mathcal{I}^A$ to $\mathcal{I}^B$

$\mathcal{I}^A$ Image list			
Query image	First rank	Second rank	Third rank
I_1^A	I_3^B	I_1^B	I_2^B
I_2^A	I_1^B	I_3^B	I_2^B
I_3^A	I_3^B	I_2^B	I_1^B

(b) From $\mathcal{I}^B$ to $\mathcal{I}^A$

$\mathcal{I}^B$ Image list			
Gallery image	First rank	Second rank	Third rank
I_1^B	I_1^A	I_2^A	I_3^A
I_2^B	I_2^A	I_3^A	I_1^A
I_3^B	I_1^A	I_2^A	I_3^A

Table II. Comparison of matching rates with other image matching and existing methods

Comparison method		VIPeR [3] (%)	CUHK01 [49] (%)	iLIDS-VID [50] (%)	PRID [51] (%)
Individual matching	SDALF [14]	19.87	—	—	—
	MSDALF	20.47	22.80	14.30	4.50
Simultaneous matching	PRiSM-II [2]	36.71	**50.10**	20.00	—
	HSV + GM	5.70	1.61	0.67	2.50
	MSDALF + GM	14.53	4.79	7.00	6.00
	HSV + HM	34.21	13.24	13.33	4.50
	MSDALF + HM	39.18	16.39	**34.40**	**16.50**
	Proposed: HSV + SMA	**40.44**	<u>27.03</u>	17.93	4.00
	Proposed: MSDALF + SMA	<u>40.32</u>	21.73	<u>34.33</u>	<u>13.00</u>

Bold numbers indicate the best matching rate result, underlined numbers indicate the second best matching rate result.

In SMA, matching will be performed by looking up both Table I(a) and (b) alternately. First, I_1^A and I_3^B are matched as I_3^B is the best choice for I_1^A, and I_1^A is also the first rank for I_3^B. Next, I_2^A is matched with I_1^B, as I_2^A is the second rank in I_1^B and it still is not matched with any image yet. In the third loop, I_3^A is matched with I_3^B, but since I_3^B is already matched with I_1^A, I_3^A remains not matched in the first loop and I_3^B is removed from the rank list of I_3^A. In the fourth loop, I_3^A which is not matched yet is matched with I_2^B. Finally, we have three matching image pairs which are (I_1^A, I_3^B), (I_2^A, I_1^B), and (I_3^A, I_2^B).

Here, we can see the query image is matched to all gallery images simultaneously by considering the similarity rank of each image. It brings us a new versatile approach compared to the conventional person re-identification methods which just implement standard individual image matching for each person image. Applying SMA in person re-identification directs us to gain a stable image matching by avoiding redundant image matches.

4. Evaluation Setting

4.1. Dataset To evaluate the performance of the proposed method, we conducted experiments with several comparison studies reported in the following subsections. For comparative study, we use the VIPeR dataset [3], CUHK01 dataset [49], iLIDS-VID dataset [50], and PRID dataset [51].

4.2. Comparative methods We compare the proposed method with individual image matching and simultaneous image matchings. For these methods, we compare two image features, HSV and MSDALF described in the previous section. We also compare with the state-of-the-art bipartite graph matching method, GM, HM, and PRiSM [2]. For the HSV feature, we used an $M = 16$-bins HSV color histogram.

4.3. Evaluation settings We follow the experimental setting described in [14] using the VIPeR dataset images as our general evaluation setting. We randomly selected 316, 485, 150, and 200 person images from VIPeR, CUHK01, iLIDS-VID, and PRID datasets, respectively, for evaluation.

In the second evaluation, we follow Zhang *et al.* [2], which focuses on a case where the number of images from two cameras are not the same. They selected the half (50.0%), one fourth (25.0%), and one eighth (12.5%) of the images in the query images set. Every query image will have only one matched image in the gallery set, but not all the gallery images are matched with images in the query set. To extend the SMA to work in an unequal number setting, we restricted the condition of the matching based on the suggestion from McVitie *et al.* [52] The proposed algorithm in phase 2 in Algorithm 1 is simply extended to the male-optimal SMA to handle the unequal numbers of images. Although Zhang *et al.* worked only on the VIPeR dataset, in this paper, we added the matching accuracy comparison with CUHK01, iLIDS-VID, and PRID datasets.

We repeated the first and second evaluations ten times while changing the selected images randomly, and then averaged the results as the final result, as similar to Farenzena *et al.* [14] Since SMA outputs only the matching result, we only considered the Rank-1 image matching score for the evaluation. For evaluating the performance of the proposed method, we calculated the Rank-1 rate from the number of successful matching over the number of persons.

We also measured the computation time for the matching by all the methods, considering Zhang *et al.*'s method as the baseline. Since implementations of the previous works are not available, we borrowed their results from their papers. Our experiments were all run on a multi-thread CPU (Intel Corei7-7700) 3.6 GHz with 16 GB of RAM.

4.4. Comparison on matching accuracy The results for the comparison on matching accuracy are summarized in Table II as two categories individual and simultaneous matching. The proposed method with a new masking, namely, MSDALF, outperformed the original SDALF for the VIPeR dataset in the individual matching comparison. This improvement from the original image masking in Farenza's work [14] how they separate the foreground and the background of the images before extracting the image feature, proved that manipulating the mask in SDALF can improve the matching rate.

In the simultaneous matching comparison, the proposed method with HSV feature outperforms other comparative methods on the VIPeR dataset. The proposed method with MSDALF feature ranked the second. Both of these proposed methods are better than PRiSM by more than 3%. In the case of the CUHK01 dataset, its images have many occlusions with other pedestrians compared to other datasets.

Furthermore, since the CUHK01 dataset images are captured at an outdoor scene with a large variety of lighting condition and changes of the illumination influence the captured images. Thus, we consider that image features were not accurately extracted in our method. For the other two datasets, iLIDS-VID and PRID, the proposed method with MSDALF ranked the second place with 0.07 and 3.50% matching accuracy behind MSDALF + HM. Through overall comparison using all the datasets, the proposed method performed well.

We consider that in general, SMA is quite promising in terms of the matching performance since it allows the image pair to be matched again; SMA allows the matched image pair to be rematched if the current similarity rank for one image is lower than that for an existing image-matched pair. However, although the proposed method with MSDALF is expected to be the best among all the comparison methods, it ranked the second. Figure 4 illustrates that Mask R-CNN could struggle in providing a promising masking image for the VIPeR dataset images.

(a)

(b)

(c)

(d)

Fig. 4. Examples of unsuccessfully masked images (using Mask R-CNN) for the four datasets used in our evaluation

Although it performed well in various situations, some of them were not well masked by Mask R-CNN due to the bad quality of images. This masking issue will influence the MSDALF feature extraction since SDALF requires these segmentation masks for splitting the image into three parts: head, torso, and legs. Mainly, some body parts such as legs are often missed in person image masks. This missing part will decrease the quality of feature extraction. As a result, the matching rate using the MSDALF feature degrades.

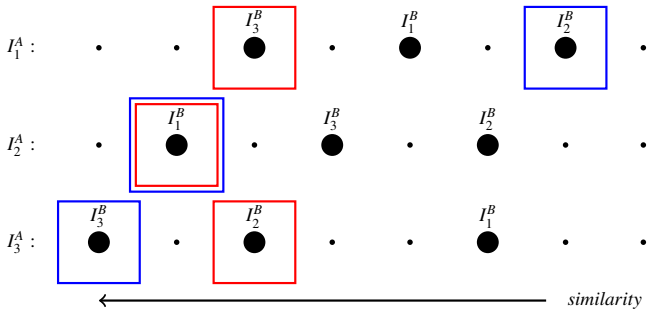

Fig. 5. Comparison of the matching results between the Greedy Matching (blue rectangular) and the proposed method utilizing stable marriage algorithm (red rectangular). The rectangulars indicate the matched images

Table III. Comparison of matching accuracy with different numbers of images in two camera views settings on the VIPeR dataset [3]

	VIPeR		
Comparison method	158 queries (%)	79 queries (%)	40 queries (%)
PRiSM-II [2]	35.90	36.70	34.50
HSV + GM	6.39	6.33	6.50
MSDALF + GM	15.00	15.06	14.75
HSV + HM	34.56	33.04	33.00
MSDALF + HM	39.49	40.00	**39.75**
Proposed: HSV + SMA	<u>40.25</u>	<u>40.20</u>	<u>39.25</u>
Proposed: MSDALF + SMA	**41.01**	**40.38**	**39.75**

Bold numbers indicate the best matching accuracy result, underlined numbers indicates the second best matching accuracy result.

For the iLIDS-VID and PRID datasets, we consider that the low color quality of the images could be the reason why the overall matching rates are not outstanding. On the other hand, for the VIPeR and CUHK01 datasets, the matching accuracy seems to be better compared to the two previous datasets. We consider that the image quality of the VIPeR dataset is acceptable for the purpose of normal eye observation, but is not acceptable for segmenting the foreground with the background.

Previous matching methods such as GM and HM are based on absolute similarity values, while the proposed method is based on similarity ranking. Figure 5 shows the difference of matching results between GM and the proposed method. In this case, by focusing only on the absolute similarity score, GM selects high similarity pairs first I_3^B for I_3^A, and I_1^B for I_2^A. Then it selects a pair for I_1^A from the leftovers I_2^B, which has low similarity scores. On the other hand, since the proposed method does not care about the absolute value, but just the ranking, it can select convincing pairs.

4.5. Comparison on different numbers of images

We show the comparison of the Rank-1 matching results on several cases where the numbers of query images are different for the VIPeR and CUHK01 datasets in Tables III and IV. The results show the robustness to the different numbers of query images of the proposed method compared to the other methods. Tables V and VI illustrate the same comparison on the iLIDS-VID and PRID datasets. The proposed method outperforms the other comparative methods for the VIPeR, CUHK01, and iLIDS-VID datasets.

The proposed method with HSV feature (HSV + SMA) ranked the first for the CUHK01 dataset, and that with MSDALF (MSDALF+SMA) also ranked the first for the VIPeR and the

Table IV. Comparison of matching accuracy with different numbers of images in two camera views settings on the CUHK01 dataset [49]

	CUHK01		
Comparison method	243 queries (%)	121 queries (%)	61 queries (%)
PRiSM-II [2]	—	—	—
HSV + GM	1.56	0.82	1.74
MSDALF + GM	4.44	4.30	3.77
HSV + HM	13.46	13.80	13.93
MSDALF + HM	16.50	16.53	16.07
Proposed: HSV + SMA	**26.46**	**28.35**	**29.18**
Proposed: MSDALF + SMA	22.63	22.40	22.30

Bold numbers indicate the best matching accuracy result, underlined numbers indicates the second best matching accuracy result.

Table V. Comparison of matching accuracy with different numbers of images in two camera views settings on the iLIDS-VID dataset [50]

	iLIDS-VID		
Comparison method	75 queries (%)	38 queries (%)	19 queries (%)
PRiSM-II [2]	—	—	—
HSV + GM	0.80	0.26	0.00
MSDALF + GM	7.33	7.89	8.42
HSV + HM	13.20	13.95	14.21
MSDALF + HM	**35.07**	**36.05**	**40.53**
Proposed: HSV + SMA	18.27	18.42	19.47
Proposed: MSDALF + SMA	**35.07**	**36.05**	38.42

Bold numbers indicate the best matching accuracy result, underlined numbers indicates the second best matching accuracy result.

Table VI. Comparison of matching accuracy with different numbers of images in two camera views settings on the PRID dataset [51]

	PRID		
Comparison method	100 queries (%)	50 queries (%)	25 queries (%)
PRiSM-II [2]	—	—	—
HSV + GM	2.00	2.80	2.80
MSDALF + GM	5.80	7.60	6.80
HSV + HM	4.70	5.20	4.80
MSDALF + HM	**16.00**	**17.20**	**15.60**
Proposed: HSV + SMA	4.30	4.80	6.40
Proposed: MSDALF + SMA	13.30	15.60	14.40

Bold numbers indicate the best matching accuracy result, underlined numbers indicates the second best matching accuracy result.

iLIDS-VID datasets. For the PRID dataset, the proposed method ranks the second after the MSDALF + HM. In total, the proposed method performs well in the case where the numbers of images are different.

4.6. Comparison of computational time The computational time during testing is the essential element in the real implementation of any system. In this paper, we compare the performance of the methods without feature extraction time and distance calculation time, and only focus on image matching time. We set the number of images to be 316 in both camera views, which is the same setting as in the VIPeR dataset. The comparison results are shown in Table VII. Although the GM is too simple to compare with the other methods, we put it in evaluation in Table VII.

Table VII. Comparison of computation time with other image matching and existing methods

Comparison method	VIPeR [3] (s)	CUHK01 [49] (s)	iLIDS-VID [50] (s)	PRID [51] (s)
PRiSM-II [2]	1.500	—	—	—
MSDALF + GM	0.002	0.004	0.001	0.001
MSDALF + HM	65.738	510.259	7.117	26.762
Proposed: MSDALF + SMA	**0.070**	**1.095**	**0.018**	**0.066**

Bold numbers indicate the best matching accuracy result.

From the overall result here, HM takes a long time for the matching. The proposed method using MSDALF + SMA showed better performance than that using MSDALF + GM. The person re-identification with not only color feature but also with other elements such as a geometrical and local features in MSDALF could gain better performance in matching. Here, we proved that the image matching scheme plays a vital role in achieving an excellent matching performance for person re-identification.

5. Conclusion

By interpreting the simultaneous person image matching problem as an instance of the marriage problem, we introduced the SMA for person re-identification. The proposed method achieved promising performance compared to all the comparative methods on the VIPeR, CUHK01, iLIDS-VID, and PRID datasets. The comparable achievements on the four datasets show that the proposed method could be a new alternative for person re-identification. The proposed method outperformed all the other comparison methods in the case where the numbers of images from two cameras are different, which is more reliable and closer to the real application. Furthermore, the proposed method achieved the fastest image matching time compared to other methods except for the GM. In addition, by avoiding the mismatch and redundant image pair with SMA, we improved the overall Rank-1 matching rate. In the future, we will consider several questions for extending the proposed method. One of those is extending the SMA to a more sophisticated one. Considering the reciprocal rank in SMA would be a choice to gain more accurate matching rate in real applications in our daily life, primarily security for tracking a targeted person in the dedicated area, e.g., airport or railway station. Second, we also will consider extending this simultaneous image matching method to multi-camera cases.

Acknowledgment

The authors would like to thank the Universiti Teknikal Malaysia Melaka (UTeM) and Ministry of Education (MOE) Malaysia for the financial support under the scholarship of Skim Latihan Akademik IPTA (SLAI). Parts of this research were supported by MEXT, Grant-in-Aid for Scientific Research (17H00745, 18K18070), and Kayamori Foundation of Informational Advancement.

References

(1) Yesil B. Watching ourselves: Video surveillance, urban space and self-responsibilization. *Cultural Studies* 2006; **20**(4–5):400–416.
(2) Zhang Z, Saligrama V. PRISM: Person reidentification via structured matching. *IEEE Transactions on Circuits and Systems for Video Technology* 2017; **27**(3):499–512.
(3) Gray D., Brenna S., and TaoH.. Evaluating appearance models for recognition, and tracking. *Proceedings of 10th IEEE Workshop on*

Performance Evaluation for Tracking and Surveillance, Vol. 3, No. 5, 2007; 41–48.
(4) Alom BMM, Das S, Islam MS. Finding the maximum matching in a bipartite graph. *DUET Journal* 2010; **1**(1):33–36.
(5) Kuhn HW. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* 1955; **2**(1–2):83–97.
(6) Hazewinkel M.. Greedy algorithm, in Encyclopedia of Mathematics, 2001. http://www.encyclopediaofmath.org/index.php?title=Greedy_algorithm&oldid=34629. Accessed June 24, 2018.
(7) Austin P. A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine* 2014; **33**(6):1057–1069.
(8) Pirsiavash H., Ramanan D., and Fowlkes C. C.. Globally-optimal greedy algorithms for tracking a variable number of objects. *Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition*, 2011; 1201–1208.
(9) Liu X., Song M., Tao D., Zhou X., Chen C., and Bu J.. Semi-supervised coupled dictionary learning for person re-identification. *Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014; 3550–3557.
(10) Cancela B., Hospedales T., and Gong S.. Open-world person re-identification by multi-label assignment inference. *Proceedings of 2014 British Machine Vision Conference*, 2014; 1–11.
(11) Bąk S. and Carr P.. One-shot metric learning for person re-identification. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017; 2990–2999.
(12) Gale D, Shapley LS. College admissions and the stability of marriage. *American Mathematical Monthly* 1962; **69**(1):9–15.
(13) Hashim N. M. Z., Kawanishi Y., Deguchi D., Ide I., and Murase H.. A preliminary study on optimizing person re-identification using stable marriage algorithm. *Proceedings of 2018 International Workshop on Frontiers of Computer Vision*, 2018; 1–6.
(14) Farenzena M., Bazzani L., Perina A., Murino V., and Cristani M.. Person re-identification by symmetry-driven accumulation of local features. *Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition*, 2010; 2360–2367.
(15) Gray D. and Tai H.. Viewpoint invariant pedestrian recognition with an ensemble of localized features. *Proceedings of 10th European Conference on Computer Vision*, 2008; 262–275.
(16) Javed O., Rasheed Z., Shafique K., and Shah M.. Tracking across multiple cameras with disjoint views. *Proceedings of 9th IEEE Conference on Computer Vision*, Vol. 2, 2003; 952–957.
(17) Wang X., Doretto G., Sebastian T., Rittscher J., and Tu P.. Shape and appearance context modeling. *Proceedings of 11th IEEE Conference on Computer Vision*, 2007; 1–8.
(18) Zhang Z., Chen Y., and Saligrama V.. A novel visual word co-occurrence model for person re-identification. *Computing Research Repository*. arXiv:1410.6532, 2014.
(19) Bąk S., Corvee E., Bremond F., and Thonnat M.. Multiple-shot human re-identification by mean Riemannian covariance grid. *Proceedings of 8th IEEE Conference on Advanced Video and Signal Based Surveillance*, 2011; 179–184.
(20) Bauml M. and Stiefelhagen R.. Evaluation of local features for person re-identification in image sequences. *Proceedings of 8th IEEE Conference on Advanced Video and Signal Based Surveillance*, 2011; 291–296.
(21) Liu C., Gong S., Loy C. C., and Lin X.. Person re-identification: What features are important? *Proceedings of 12th European Conference on Computer Vision*, 2012; 391–401
(22) Zhao R., Ouyang W., and Wang X.. Unsupervised salience learning for person re-identification. *Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013; 3586–3593.
(23) Cai Y. and Pietikäinen M.. Person re-identification based on global color context. *Proceedings of 10th Asian Conference on Computer Vision*, 2010; 205–215.
(24) Wu S., Chen Y. C., Li X., Wu A. C., You J. J., and Zheng W. S.. An enhanced deep feature representation for person re-identification. *Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision*, 2016; 1–8.
(25) Matsukawa T. and Suzuki E.. Person re-identification using CNN features learned from combination of attributes. *Proceedings of 23rd IAPR International Conference on Pattern Recognition*, 2016; 2428–2433.
(26) Matsukawa T., Okabe T., Suzuki E., and Sato Y.. Hierarchical Gaussian descriptor for person re-identification. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016; 1363–1372.
(27) Matsukawa T, Okabe T, Suzuki E, Sato Y. Hierarchical gaussian descriptors with application to person re-identification. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 2019; **4**:1–14.
(28) Karanam S., Li Y., and Radke R. J.. Person re-identification with discriminatively trained viewpoint invariant dictionaries. *Proceedings of 15th IEEE Conference on Computer Vision*, 2015; 4516–4524.
(29) Wang F., Zuo W., Lin L., Zhang D., and Zhang L.. Joint learning of single-image and cross-image representations for person re-identification. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016; 1288–1296.
(30) Hirzer M., Roth P. M., and Bischof H.. Person re-identification by efficient impostor-based metric learning. *Proceedings of 9th IEEE Conference on Advanced Video and Signal-Based Surveillance*, 2012; 203–208.
(31) Koestinger M., Hirzer M., Wohlhart P., Roth P. M., and Bischof H.. Large scale metric learning from equivalence constraints. *Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012; 2288–2295.
(32) Xiong F., Gou M., Camps O., and Sznaier M.. Person re-identification using kernel-based metric learning methods. *Proceedings of 13th European Conf. on Computer Vision*, 2014; 1–16.
(33) Tao D, Jin L, Wang Y, Yuan Y, Li X. Person re-identification by regularized smoothing kiss metric learning. *IEEE Transactions on Circuits and Systems for Video Technology* 2013; **23**(10):1675–1685.
(34) Paisitkriangkrai S., Shen C., and Hengel A. V. D.. Learning to rank in person re-identification with metric ensembles. *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015; 1846–1855.
(35) Li W., Wu Y., and Li J.. Re-identification by neighborhood structure metric learning. *Pattern Recognition*, Vol. 61, 2017; 327–338.
(36) Sun C, Wang D, Lu H. Person re-identification via distance metric learning with latent variables. *IEEE Transactions on Image Processing* 2017; **26**(1):23–34.
(37) Song J., Yang Y., Song Y. Z., Xiang T., and Hospedales T. M.. Generalizable Person re-identification by domain-invariant mapping network. *Proceedings 2019 IEEE Conference on Computer Vision and Pattern Recognition*, 2019; 719–728.
(38) Mao C, Li Y, Zhang Z, Zhang Y, Li X. Pyramid person matching network for person re-identification. *Machine Learning Research* 2017; **77**:487–497.
(39) Ahmed E., Jones M., and Marks T. K.. An improved deep learning architecture for person re-identification. *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*, 2015; 3908–3916.
(40) Liu J., Zha Z., Tian Q., Liu D., Yao T., Ling Q., and Mei T.. Multi-scale triplet CNN for person re-identification. *Proceedings of 24th ACM Multimedia Conference*, 2016; 192–196.
(41) Zhang L., Xiang T., and Gong S.. Learning a discriminative null space for person re-identification. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016; 1239–1248.
(42) Shen Y., Lin W., Yan J., Xu M., Wu J., and Wang J.. Person re-identification with correspondence structure learning. *Proceedings of 15th IEEE Conference on Computer Vision*, 2015; 3200–3208.
(43) Ye M., Ma A. J., Zheng L., Li J., and Yuen P. C.. Dynamic label graph matching for unsupervised video re-identification. *Proceedings of 16th IEEE Conference on Computer Vision*, 2017; 5142–5150.
(44) Manlove DF. Hospitals/residents problem: 1962; Gale, Shapley. In *Encyclopedia of Algorithms*; 2008; 390–394. Springer: New York.
(45) Sural >S., Qian G., and Pramanik S.. Segmentation and histogram generation using the HSV color space for image retrieval. *Proceedings of 2002 IEEE International Conference on Image Processing*, Vol. 2, 2002; 589–592.
(46) Barla A., Odone F., and Verri A.. Histogram intersection kernel for image classification. *Proceedings of 2003 IEEE International Conference on Image Processing*, Vol. 3, 2003; 513–516.
(47) Jojic N., Perina A., Cristani M., Murino V., and Frey B.. Stel component analysis: Modeling spatial correlations in image class

structure. *Proceedings of 2009 IEEE Conf. on Computer Vision and Pattern Recognition*, 2009; 2044–2051.
(48) He K., Gkioxari G., Dollăr P., and Girshick R.. Mask R-CNN. *Proceedings of 16th IEEE International Conference on Computer Vision*, 2017; 2961–2969.
(49) Li W., Zhao R., and Wang X.. Human reidentification with transferred metric learning. *Proceedings of 11th Asian Conference on Computer Vision*, 2012; 31–44.
(50) Wang T., Gong S., Zhu X., and Wang S.. Person re-identification by video ranking. *Proceedings of 13th European Conference on Computer Vision*, 2014; 688–703.
(51) Hirzer M., Beleznai C., Roth P. M., and Bischof H.. Person re-identification by descriptive and discriminative classification. *Proceedings of 17th Scandinavian Conference on Image Analysis*, 2011; 91–102.
(52) McVitie DG, Wilson LB. Stable marriage assignment for unequal sets. *BIT Numerical Mathematics* 1970; **10**(3):295–309.

Nik Mohd Zarifie Hashim (Non-member) received the B.E. and M.E. degrees in Electrical and Electronics Engineering from University of Fukui, Japan in 2006 and 2008. In 2007, he joined Universiti Teknikal Malaysia Melaka as a Tutor. From 2008, he became Lecturer in the same university. He is member of IEEE, Board of Engineering Malaysia (BEM), and Graduate Engineer of The Institution of Engineers, Malaysia (IEM). He is currently studying his Ph.D. in Media Science at Graduate School of Information Science in Nagoya University. His research focuses on person re-identification and 3D object pose estimation.

Yasutomo Kawanishi (Non-member) received his B.E. and M.E. degrees in Engineering and a Ph.D. degree in Informatics from Kyoto University, Japan, in 2006,2008, and 2012, respectively. He became a Post Doctoral Fellow at Kyoto University, Japan in 2012. He moved to Nagoya University, Japan as an Designated Assistant Professor in 2014. Since 2015, he has been an Assistant Professor at Nagoya University, Japan. His research interests are Pedestrian-centric Vision, which includes Pedestrian Detection, Tracking, and Retrieval, for surveillance and in-vehicle videos. He received the best paper award from SPC2009, and Young Researcher Award from IEEE ITS Society Nagoya Chapter. He is a member of IEEE, IIEEJ, and IEICE.

Daisuke Deguchi (Non-member) received his B.E. and M.E. degrees in Engineering and a Ph.D. degree in Information Science from Nagoya University, Japan, in 2001, 2003, and 2006, respectively. He became a Post Doctoral Fellow at Nagoya University, Japan in 2006. From 2008 to 2012, he had been an Assistant Professor at the Graduate School of Information Science, Nagoya University. From 2012, He had been an Associate Professor in Information Strategy Office, Nagoya University, Japan. He is working on the object detection, segmentation, recognition from videos, and their applications to ITS technologies, such as detection and recognition of traffic signs. He is a member of IEEE, IEICE, and IPS Japan.

Ichiro Ide (Non-member) received his B.E., M.E., and Ph.D. from The University of Tokyo in 1994, 1996, and 2000, respectively. He became an Assistant Professor at the National Institute of Informatics, Japan in 2000. Since 2004, he has been an Associate Professor at Nagoya University. He was also a Visiting Associate Professor at National Institute of Informatics from 2004 to 2010, an Invited Professor at Institut de Recherche en Informartique et Systèmes Aláeatoires (IRISA), France in 2005, 2006, and 2007, a Senior Visiting Researcher at ISLA, Instituut voor Informatica, Universiteit van Amsterdam from 2010 to 2011. His research interest ranges from the analysis and indexing to retargeting of multimedia contents, especially in large-scale broadcast video archives, mostly on news, cooking, and sports contents. He is a senior member of IEICE and IPS Japan, and a member of IEEE, ACM, ITE, and JSAI.

Hiroshi Murase (Non-member) received his B.E., M.E., and Ph.D. degrees in Electrical Engineering from Nagoya University, Japan. In 1980, he joined the Nippon Telegraph and Telephone Corporation (NTT). From 1992 to 1993, he was a visiting research scientist at Columbia University, New York. From 2003, he is a professor of Nagoya University, Japan. He was awarded the IEICE Shinohara Award in 1986, the Telecom System Award in 1992, the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Best Paper Award in 1994, the IPS Japan Yamashita Award in 1995, the IEEE International Conference on Robotics and Automation (ICRA) Best Video Award in 1996, the Takayanagi Memorial Award in 2001, the IEICE Achievement Award in 2002, and the Ministry Award from the Ministry of Education, Culture, Sports, Science, and Technology in 2003. He is a Fellow of IEEE, IEICE, and IPS Japan.