

Person re-identification with mutual re-ranking

Ngoc-Bao Nguyen¹ · Vu-Hoang Nguyen¹ ·
Thanh Duc Ngo¹ · Khang M. T. T. Nguyen¹

Received: 1 April 2016 / Accepted: 28 December 2016 / Published online: 19 January 2017
© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract Person re-identification is the problem of identifying people moving across cameras. Traditional approaches deal with this problem by pair-wise matching images recorded from two different cameras. A person in the second camera is identified by comparing his image with images in the first camera, independently of other persons in the second camera. In reality, there are many situations in which multiple persons appear concurrently in the second camera. In this paper, we propose a method for post-processing re-identification results. The idea is to utilize information of co-occurrence persons for comparing and re-arranging given ranked lists. Experiments conducted on different datasets with several state-of-the-art methods have shown the effectiveness of our post-processing method in improving re-identification accuracy.

Keywords Person re-identification · Ranked list · Re-ranking · Cumulative matching characteristic

1 Introduction

With the popularity of surveillance cameras, security observation systems are applied ubiquitously, especially in public

places such as supermarkets, airports, and hospitals. Such a system includes multiple cameras connected to an operation center. Operators who are displayed images recorded from cameras have to observe and perform various tasks: detecting, recognizing, and keeping track of characters. Among these tasks, tracking people crossing multiple cameras plays an important role.

This task becomes much more challenging when the number of cameras increases and there are more people appearing in camera's view. Automatic systems which can automatically recognize people across multiple cameras are needed. The essential problem of such system has recently been studied and named person re-identification. In other words, it is defined as the problem of matching human images recorded from multiple cameras distributed over non-overlapped areas, in the context of the persons crossing through many cameras.

Formally, the person re-identification problem can be formulated as follows: Given n persons crossing camera 1, some of them appear later in camera 2. For each image (or person) recorded from the camera 2 (called probe image), determine a list of images (or persons) recorded from camera 1 (called gallery images). Gallery images in the list are ranked by their likelihood of being the same person of the currently considered as probe image (see Fig. 1).

A person re-identification system receives images or videos from multiple cameras as input and output the matching of images of people appearing in those images or videos [1].

Due to the low resolution of surveillance cameras, traditional recognition methods such as biometric cues or face and iris recognition could not be applied. In addition, variation of viewpoints and illumination across different cameras, which cause appearance changes, is among the most challenging problems leading to mismatching. Other challenging issues

✉ Ngoc-Bao Nguyen
ngocntb@uit.edu.vn

Vu-Hoang Nguyen
vunh@uit.edu.vn

Thanh Duc Ngo
thanhd@uit.edu.vn

Khang M. T. T. Nguyen
khangntm@uit.edu.vn

¹ Multimedia Communications Laboratory, University of Information Technology, VNU-HCM, Ho Chi Minh, Vietnam

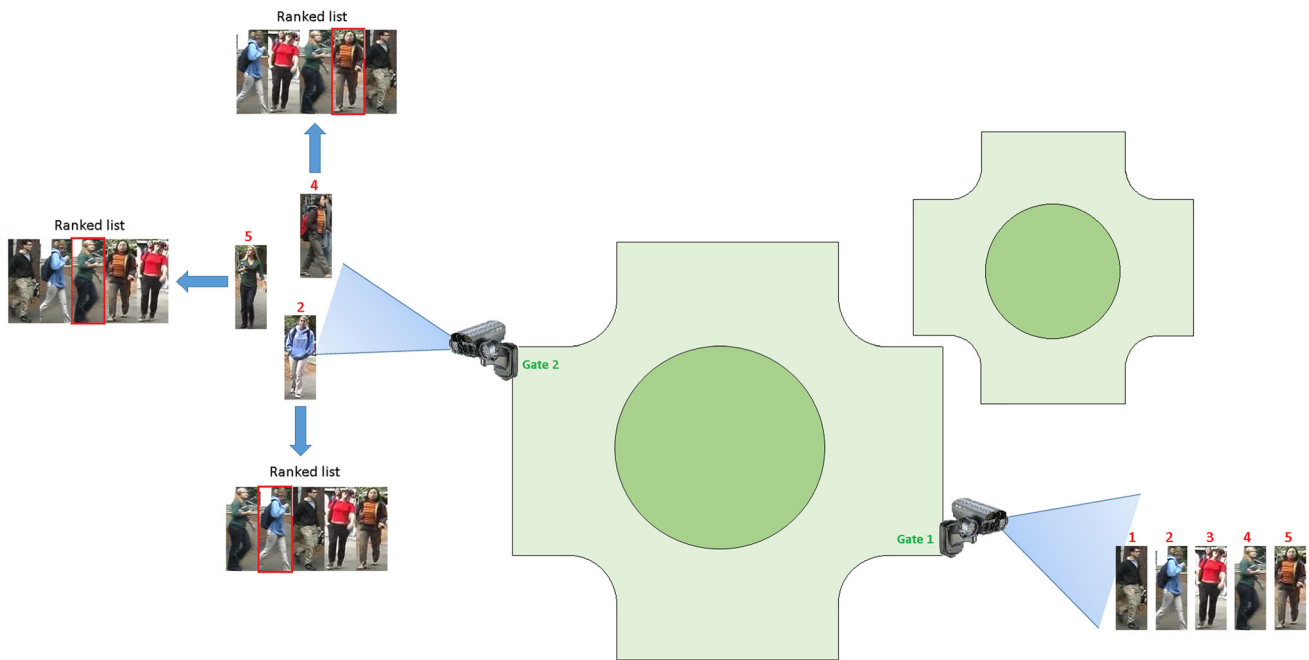


Fig. 1 An example of person re-identification with 2 cameras set up at 2 gates of the building. In this example, there are 5 people going through gate 1 under the view of camera 1. Three of them appear later

in camera 2. For each human image captured by camera 2, a ranked list of the 5 images captured by camera 1 is produced. The ground truth in each list is bordered by a *red rectangle* (color figure online)

relating to person re-identification are occlusion and background clutter.

A typical person re-identification pipeline consists of two components: feature extraction and image matching. State-of-the-art methods usually employ multiple features. SDALF [6] integrates three kinds of low-level features: weighted color histogram, maximally stable color regions (MSCR), and recurrent high-structured patches (RHSP). Meanwhile, semantic features are used in [13] together with other low-level features. Another approach is to learn appropriate metrics for specific data [25].

With existing person re-identification systems, probe images of individual persons are treated independently. Given a probe image, they compute the distances from gallery images to the probe image. Using the computed distances, a ranked list of gallery images is then generated.

However, in reality, there are cases in which multiple persons appear concurrently in a camera. Human beings could assess and utilize such information to give a more accurate prediction. Namely, if a gallery image of one person is ranked very high in a list, that image should be ranked very low in other lists, given the ranked lists of different probe images. In this paper, we propose a method using such constraint in post-processing to improve identification accuracy. Information of co-occurrence persons is employed to mutually re-rank the returned lists. Specifically, each highly ranked gallery image in a list is assigned a penalty. The penalties are then used to

update scores of the gallery images in other lists. We compute the penalties based on similarities of gallery images to probe images.

Compared to existing work [23], we provide two main extensions:

- First, we study the generality of the proposed approach with different penalty functions. We evaluate two penalty functions (i.e. Penalty I and Penalty II). Using two functions presenting the idea in different ways, we learned that both functions helped to improve the performance of the original person re-identification method.
- Second, more experiments were conducted. In [23], only one person re-identification method, SDALF [6], is evaluated on VIPeR [10]. In this work, we extensively consider four state-of-the-art person re-identification methods including SDALF [6], QAF[33], Mid-level filter [32], and SDC_{knn} [30]. These methods were evaluated on three different benchmark datasets: VIPeR [10], ETHZ [5,26], and CUHK01 [15]. By doing this, we expect to provide a comprehensive evaluation of the proposed approach.

The remainder of this paper is organized as follows: Sect. 2 is an overview of related works. Section 3 presents our proposed re-ranking method. Experimental results are shown in Sect. 4. Finally, Sect. 5 is the conclusion of the paper.

2 Related works

There are two main parts in a typical person re-identification system: feature extraction and similarity estimation. Re-ranking is usually applied in the post-processing process to improve identification accuracy.

Low-level features are widely used in feature extraction. In [6], weighted histogram and blobs are used. Specifically, color histograms are computed with weights. The weights are based on distances of pixels to the asymmetric axis of the person image. Besides, the authors extracted blobs using the method called maximally stable colour regions (MSCR) in [7]. In [24], the authors detected blobs on person images. Then, they extracted color histogram and histogram of oriented gradient (HOG) for visual features. Ma et al. [21] introduced a new feature, called biologically inspired feature (BIF). It is extracted by convolving images with Gabor filters. Then, MAX pooling is applied for two convolved images with consecutive bands. Prosser et al. [25] and Gray et al. [11] focused on color and texture features. Specifically, 8 color channels from RGB, HS, and YCbCr color systems are used. The authors used Gabor and Schmid filters on the luminance channel for texture features. In [33], the authors used local features and proposed an unsupervised method for determining feature weight for fusion. Local descriptors of pixels are transferred into Fisher Vectors to represent images in [22]. Unlike other image retrieval problems, local features are not commonly used in person re-identification [9, 12].

Mid-level features built from low-level features are used in recent study due to their high-level abstraction and efficiency. In [32], selected discriminative and representative local patches are used for learning mid-level feature filters. In [16], the authors used a deep learning framework to learn pairs of mid-level filters which encode the transformation of mid-level appearance between the two cameras. Inspired by the recognition ability of human being, the authors in [31] proposed an unsupervised method for detecting salient and distinctive local patches and used them for matching images.

Semantic features, human understandable mid-level features, are applied in [13, 20] for person re-identification. In [13], semantic features are first detected by applying SVM with texture features (introduced in [25]). The detected mid-level features are then used for re-identification. Liu et al. proposed to use topic models [20] to represent the attributes (i.e. semantic mid-level features) of images.

Feature selection and weighting were also addressed in recent related works. In [19], the authors used an unsupervised approach to adaptively identify the confidence of features in different circumstances. Or in [11], the authors defined a feature space. Then, they proposed a learning approach to search for the optimal representation. Zhao et al. [30] focused on extracting discriminative image patches and learning human salient descriptor from them.

To accurately match person images, body part localization is required. The method, named SDALF in [6], employed a simple body part detector. The aim was to determine upper and lower parts of a person image by a horizontal line. The line is set so that the separated parts have minimum difference in area and maximum difference in color. More complicatedly, Gheissari et al. [9] used a technique, called decomposable triangulated graph, for localizing human body parts. Triangulated graphs are fitted to human images by minimizing the energy function. Given the fitted graph, body parts can be localized for matching. Or in [2], pictorial structures were applied for detecting body configuration.

Given extracted features, similarity estimation is also important to a person re-identification system. Besides traditional distances like L1 distance, L2 distance, or Bhattacharyya distance, recent works also focus on learning a new type of distance [25, 26]. In Prosser et al. [25] reformulated person re-identification as a ranking problem in which they learn a ranking function. The function ranks relevant pairs higher than irrelevant pairs. In Schwartz and Davis [26], the authors use partial least square to learn the weights of different features by considering their discrimination. The other direction is to learn the transformation between two cameras. In Zhao et al. [29], a function is defined to present the transformation from a fixed camera to another fixed camera. Zheng et al. [34] considered person re-identification as a relative distance comparison learning problem which aims at learning appropriate distances for each pair of images. Different from other works, Zhen et al. [17] proposed to simultaneously distance metrics and the decision threshold instead of the distance metrics only.

In general, re-ranking approaches for image retrieval can be applied to person re-identification. In [28], users' intentions expressed in their feedback are used to re-rank the output lists. In [4], the authors proposed a method for query expansion by using images selected from the initial ranked list. Similarly, top images from an output ranked list are used for re-querying [3]. By doing this, more relevant images are returned. Assuming relevant images are highly similar to the nearest neighbors of a query, the authors in [27] introduced a method to accurately localize interest instances in the retrieved images. The features extracted from localized instances in top ranked images are then used to refine the retrieval results.

There are several recently proposed re-ranking methods dedicated to person re-identification. The authors in [14] claimed that true matched pairs of images are supposed to have many common visually similar neighbors, called context similarity, in addition to mutual visually similar appearance, called content similarity. They suggested reversely querying each gallery image with newly formed set including the probe image and other gallery images. Then, the initial result is revised using the bidirectional ranking

lists. Inspired by [14], Garcia et al. [8] proposed to eliminate ambiguous cases in the first ranks of the lists by assuming that the ground truths appear in the first ranks of the lists as well. Unlike the above two works which utilize internal information of ranked lists for optimization, the authors in [18] designed an interactive method which allows users to pick strong and weak negative samples from the returned list. The selected negative samples will then be used to refine the list. This approach is very important in practical applications when users need acceptably accurate results.

In this paper, we introduce a method to improve person re-identification accuracy by utilizing information of the co-occurrence of people for re-ranking. To the best of our knowledge, such information has not been explicitly employed in existing person re-identification approaches.

3 Re-ranking with co-occurrence constraints

In this section, we introduce our proposed post-processing method. Given initial ranked lists returned by a person re-identification system, our method is then used to re-rank the lists by taking co-occurrence constraints into account.

3.1 Definitions

The traditional person re-identification problem can be stated as follows:

Given n persons crossing camera 1, their images are captured to generate an image set, named gallery images. There is a person crossing camera 2 and his image is called probe image. The task is to return a list of gallery images of being the same person as the probe image.

Existing person re-identification methods treat probe persons independently of each other. However, in real applications, we learn that there are cases in which multiple persons appear at the same time and within the camera's observation regions.

Figure 2 presents a scenario which two persons co-occur in the same camera and their results of re-identification. The numbers in brackets represent probabilities of the persons in the probe image and the gallery image are the same person. The probabilities are defined based on their similarity scores. Here, we assume two probe persons (Probe 1 and Probe 2) co-occur in Camera 2. With the first probe image (Probe 1), the image X is significantly more similar to the probe image than other gallery images of the list, according to their similarity scores. Hence, X can be considered as a correct match. Whereas, with the second probe image (Probe 2), because their similarity scores are slightly different, it is difficult to identify the correct one. However, if the information from the first rank list is provided, i.e. X is Probe 1, we can refine the list by moving X toward the end of the second ranked list. In other words, this means if X is more likely to be Probe 1,



Fig. 2 An example of re-ranking: **a** Probe image 1 and its ranked list; **b** Probe image 2 and its ranked list; **c** Probe image 2 and its re-ranked list based on ranked list (a)

it should not be Probe 2 at the same time. By doing this, we may pull correct match to a lower rank (i.e. closer to rank 1) while pushing the incorrect match to a higher rank (as shown in c). As a result, the accuracy is improved.

Inspired by such observation, our proposal is to co-occurrence constraints of multiple probe persons to refine ranked lists initially returned by a person re-identification method for a higher accuracy (see Fig. 3). In such a context, our re-ranking problem can be stated as follows:

- *Assumption* There are multiple probe persons appearing concurrently.

- *Input* Ranked lists of those probe persons initially generated by a person re-identification method.
- *Output* re-ranked lists with higher accuracy.

3.2 Re-ranking method

Here, we describe the proposed re-ranking method in detail.

Assuming that we have k probe persons appearing at the same time and n gallery persons, using a person re-identification method, k ranked lists and scores of the gallery images in each list (higher score means higher distance to the probe image, and thus, less similar to the probe image) are obtained. The more similar to probe image a gallery image is, the higher rank it should be in other lists of probe images. Therefore, we introduce a penalty score computed for each gallery image with respect to each ranked list. Scores of gallery images in each list are updated using penalties and the lists are rearranged according to new scores.

The penalty score of each gallery image with respect to each ranked list can be computed from the distance of that image to the probe image of the ranked list by using penalty functions. With those functions, the more different to the probe image a gallery is, the lower penalty it will receive from the corresponding ranked list. In this paper, we propose two penalty functions which we call Penalty I and Penalty II. However, it is worth noting that any other functions with the property discussed above can be applied, independent of the method for person re-identification.

Penalty I:

$$penalty(Img_i, L_j) = e^{-distance^2(Img_i, Probe_j)/\gamma^2} \tag{1}$$

Penalty II:

$$penalty(Img_i, L_j) = \frac{1}{1 + e^{distance^2(Img_i, Probe_j)/\beta^2}}, \tag{2}$$

where Img_i is the i th gallery image. $Probe_j$ is the j th probe image, and L_j is its corresponding ranked list. The distance function indicates the confidence score of being the same person of two images. That score is initially returned by the person re-identification method. γ and β are parameters to control the variance of penalties.

Gallery images in the initial lists are ranked by their confidence scores with probe images. In this paper, the scores in one list are updated using penalties computed from other lists.

$$newscore(Img_i, L_j) = originalscore(Img_i, L_j) + \frac{1}{k-1} \sum_{q \neq j} penalty(Img_i, L_q), \tag{3}$$

where $originalscore$ and $newscore$ are, respectively, the original distance and updated distance between gallery images and the probe image of the list, Img_i is the i th gallery image, L_j is the j th list, and k is the number of people appearing at the same time. A large penalty of a gallery image in a list will increase the distance of that image to the probe images in other lists.

The final ranked lists are produced by sorting images based on their new scores.

RE-RANKING ALGORITHM

Input k ranked lists of k co-occurrence persons

Output k re-ranked lists with higher accuracy (*expected*)

```

for  $i = 1 \rightarrow length(L_i)$  do
  for  $j = 1 \rightarrow k$  do
    | compute  $penalty(Img_i, L_j)$ 
  end
end
for  $i = 1 \rightarrow length(L_i)$  do
  for  $j = 1 \rightarrow k$  do
    |  $newscore(Img_i, L_j) = originalScore(Img_i, L_j) +$ 
    |  $1/(k-1) \times \sum_{q \neq j} penalty(Img_i, L_q)$ 
  end
   $L_i = sort(newscore(i));$ 
end

```

4 Experiments

4.1 Experimental settings

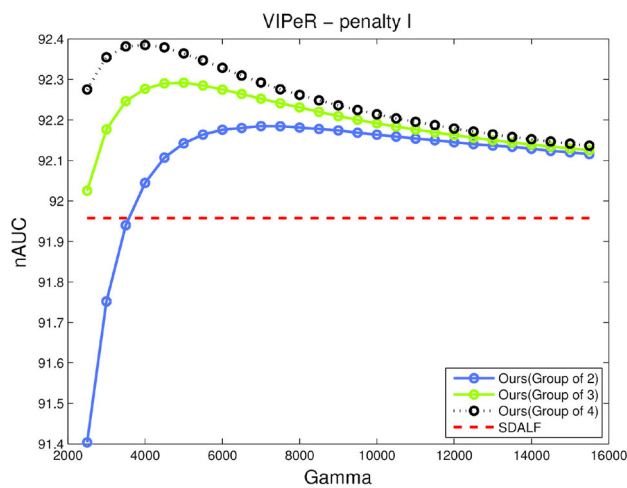
To evaluate and compare performances of different methods, Cumulative Matching Characteristic (CMC) is widely used. CMC [10] represents the frequency of the correct match standing in top n of the ranked list. Specifically, a point (x, y) in the curve means that there is $y\%$ of the lists having ground truth in top x . Accordingly, the higher curves represent the more accurate lists. However, if the curves of different methods are not much distinctive to each other, it is not easy to compare them. We, therefore, employ area under curve (AUC) scores for the CMC curves. AUC score is the area bounded between by the curve and the x -axis. Higher values of AUC indicate better performance. AUC scores are typically normalized so that the highest AUC will be 100. Normalized AUC (nAUC) is used in this paper for evaluation.

In order to verify the effectiveness of the proposed re-ranking method, we select 4 state-of-the-art person re-identification methods: SDALF [6], MidFilter [32], Query Adaptive late Fusion (QAF) [33], and SDC_{knn} [30] for experiments. Given initially ranked lists returned by those methods, we then apply the proposed re-ranking method to the lists.

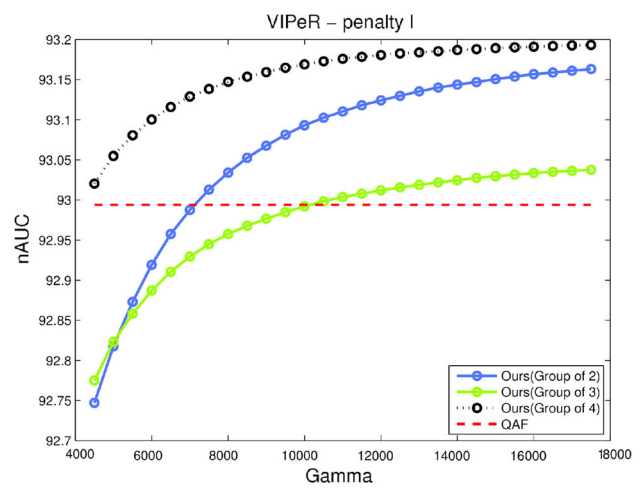
SDALF [6] With this method, each human body image is divided into upper part and lower part by a horizontal line. The line is tuned to maximize the color dissimilarity



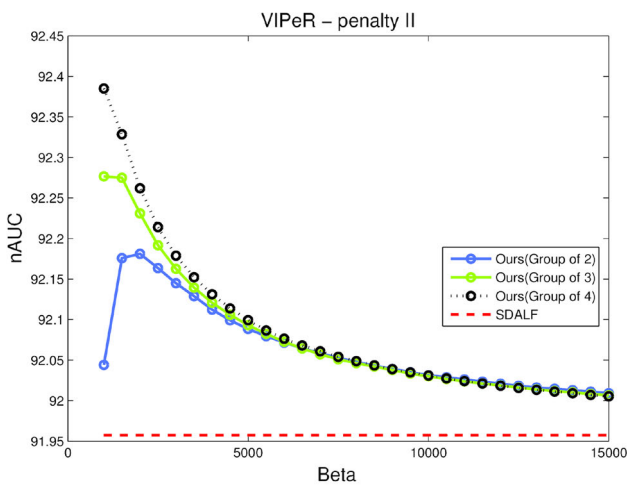
Fig. 3 Re-ranking for re-identification with the context of k probe persons appearing simultaneously



(a) Penalty I

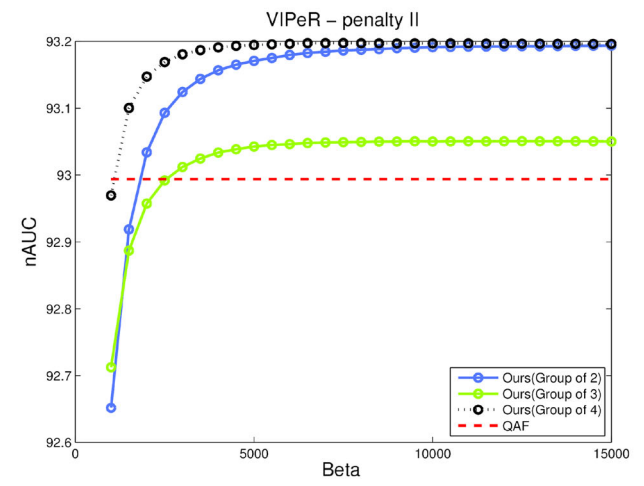


(a) Penalty I



(b) Penalty II

Fig. 4 nAUC scores of the SDALF method and re-ranking method with different γ and β on VIPeR



(b) Penalty II

Fig. 5 nAUC scores of the Query-Adaptive Late Fusion method and re-ranking method with different γ and β on VIPeR

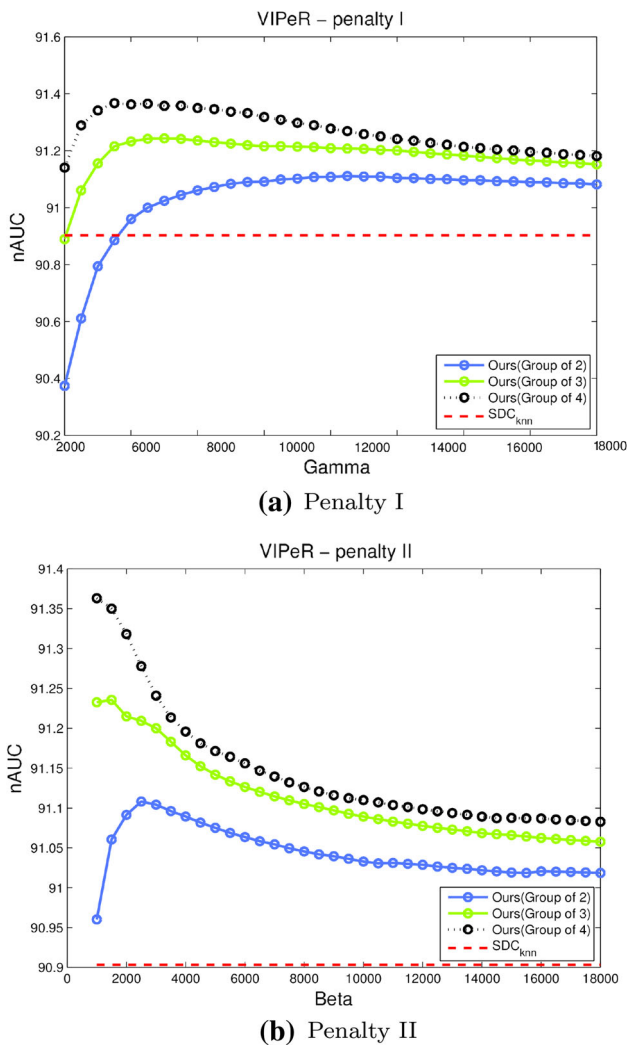


Fig. 6 nAUC scores of the SDC_{knn} method and re-ranking method with different γ and β on VIPeR

and minimize the area difference between the two parts. Different types of visual features such as weighted histogram, maximally stable colour regions (MSCR) [7], and Recurrent High-Structured Patches (RHSP) are then extracted on each part.

MidFilter [32] Unlike [6], which relies on low level features, the method in [32] focuses on learning mid-level patches for representing human images. Image patches are collected from the image set, qualified into discriminative and representative scores, hierarchically clustered. The patches which are both discriminative and representative are kept for image representation.

SDC_{knn} [30] In this method, Zhao et al. claim that humans can easily distinguish people by identifying their discriminative features. Hence, they design a method to extract salient features of pedestrian images. Salient patches are then used to learn a human salient descriptor for images in an unsupervised manner.

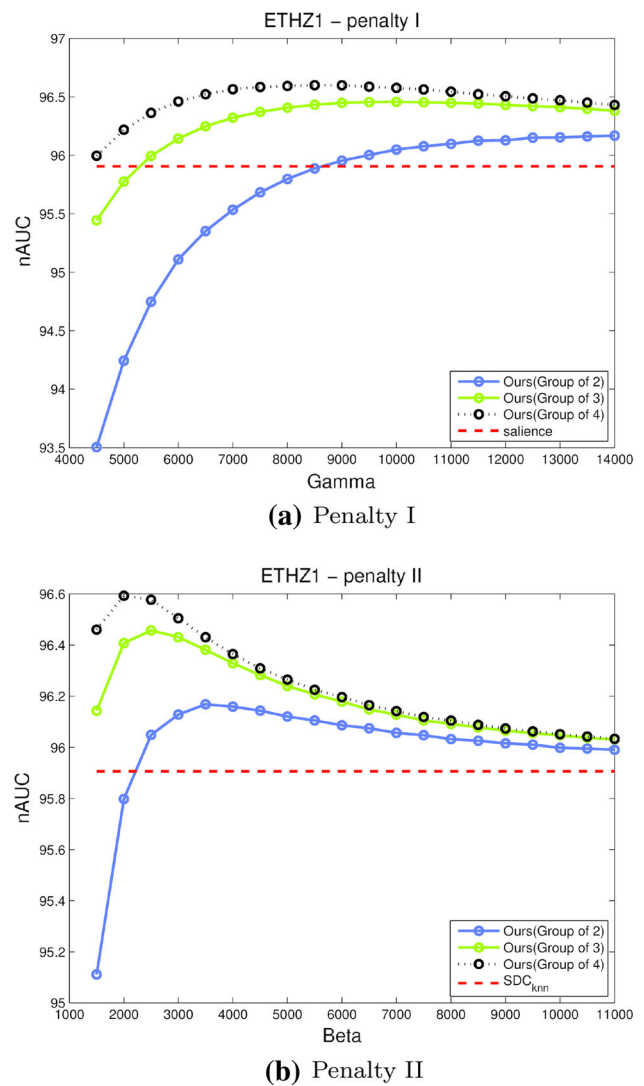
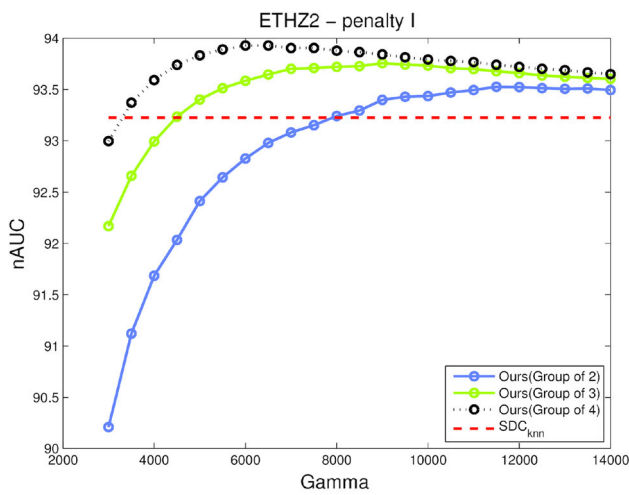


Fig. 7 nAUC scores of the SDC_{knn} method and re-ranking method with different γ and β on ETHZ1

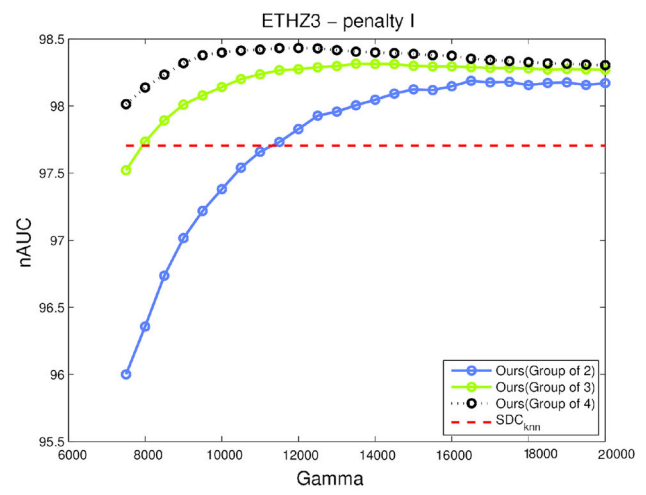
QAF [33] The authors focus on estimating weights for different features adaptively with each query or probe image. More specifically, based on the shape of the score list of each feature type when querying, the method can estimate the effect of the feature, determining its weight for fusion. The method uses local features including H-S histograms, Color Names, LBP, and HOG together with Bag-Of-Words (BoW) model.

We conduct experiments on benchmark databases including VIPeR [10], ETHZ [5,26], and CUHK01 [15].

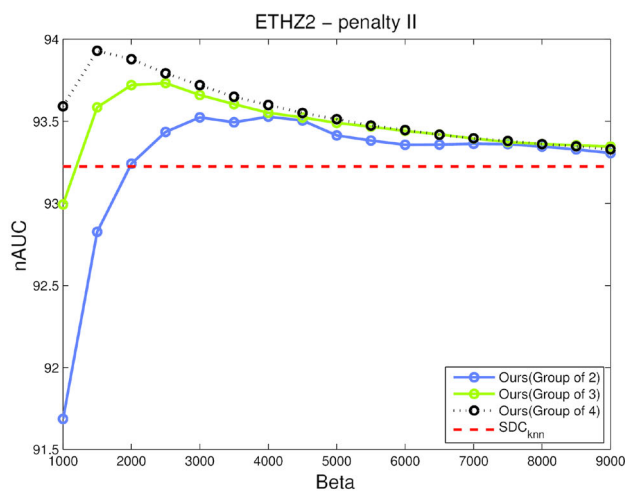
VIPeR [10] (Viewpoint Invariant Pedestrian Recognition) is a standard dataset for person re-identification problem and is considered as one of the most difficult datasets. VIPeR contains 1264 images of 632 pedestrians. Each pedestrian is represented by two images from different cameras. The challenges of this dataset are viewpoint changes (around



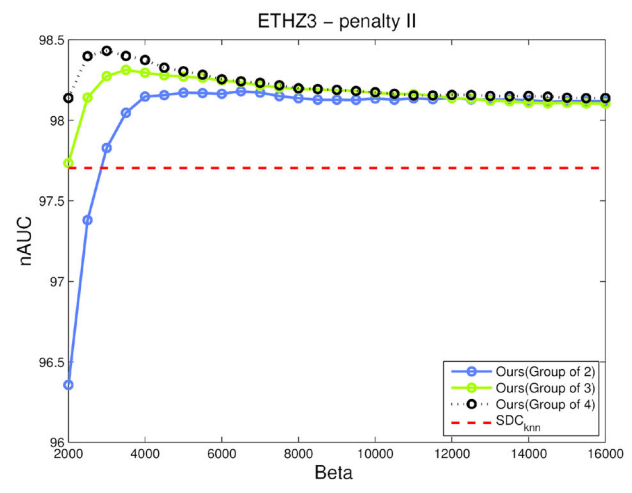
(a) Penalty I



(a) Penalty I



(b) Penalty II



(b) Penalty II

Fig. 8 nAUC scores of the SDC_{knn} method and re-ranking method with different γ and β on ETHZ2

Fig. 9 nAUC scores of the SDC_{knn} method and re-ranking method with different γ and β on ETHZ3

90 degrees for most of pairs of images) and illumination changes. Besides, low resolution of images in VIPeR is also a factor degrading performances significantly. In this dataset, each pair of images is divided into two sets, CamA and CamB. CamA and CamB are then considered as gallery set and probe set or vice versa. The VIPeR dataset is used with SDALF, QAF, and SDC_{knn} with similar settings as in the papers.

The **ETHZ** dataset [5] consists of 3 subsets: ETHZ1, ETHZ2, ETHZ3. Each subset is recorded from a camera stuck on a moving wagon. Schwartz and Davis [26] have applied person detection on the ETHZ subsets to crop human images from the raw video. After detection, ETHZ1 contains 4857 images of 83 characters. ETHZ2 and ETHZ3 include 1936 and 1762 images of 35 and 28 persons respectively. In the ETHZ datasets, we randomly choose a pair of images for each person. Half of them are considered as gallery images,

the remaining is considered as probe images. The ETHZ dataset is used with the SDALF and SDC_{knn} method.

CUHK01 [15] consists of front view and back view images of 972 people which are used as gallery and probe images in the experiment. The images in CUHK01 are resized to 160×60 for standardization. CUHK01 is used for experiments of Mid-level Filters with the similar setting in the paper.

In order to re-rank, we need the information of multiple probe people appearing concurrently. This kind of information is not available in person re-identification datasets. Therefore, we simulate such cases by randomly clustering images of each dataset into groups of k persons. Within each group, we have k ranked lists corresponding to k probe persons appearing concurrently. In each group, the lists are then mutually re-ranked by the proposed method. In this experiment, we try with groups (also called batch) of two, three,

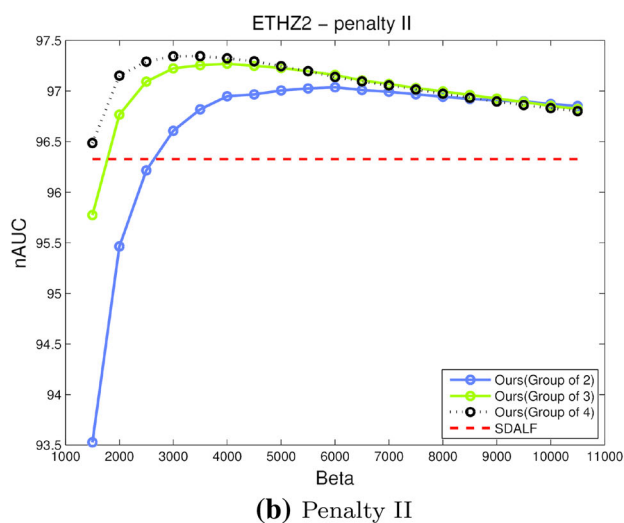
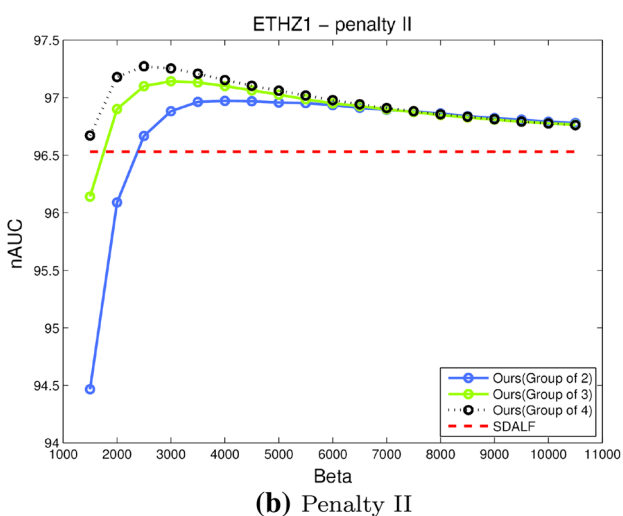
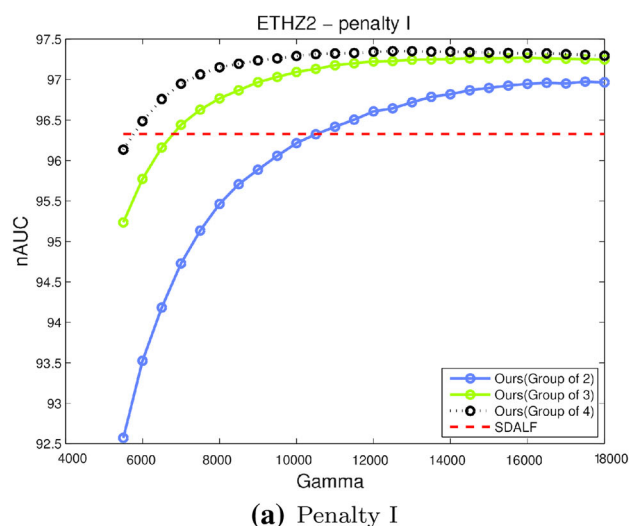
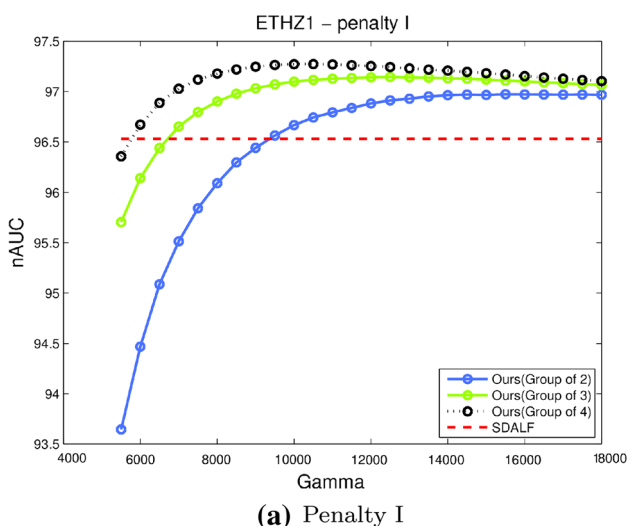


Fig. 10 nAUC scores of the SDALF method and re-ranking method with different γ and β on ETHZ1

Fig. 11 nAUC scores of the SDALF method and re-ranking method with different γ and β on ETHZ2

and four persons. Both types of penalty function are applied to the experiments. Because the performance of our method depends on each permutation of groups, we repeat the experiments 200 times and take the average result.

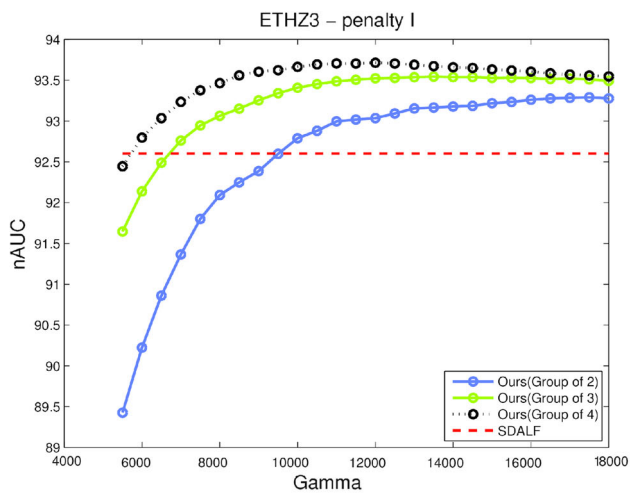
4.2 Results and analysis

The results when applying our method on SDALF and QAF on VIPeR are shown in Figs. 4, 5, and 6. Overall, we learn that the person re-identification accuracy is improve after the re-ranking process. For SDALF and SDC_{knn} , nAUC is increased up to approximately 0.5. 0.2 nAUC improvement is made for QAF method on VIPeR dataset. An interesting point to notice is that re-ranking in groups of four improves the performance the most in all of the three methods.

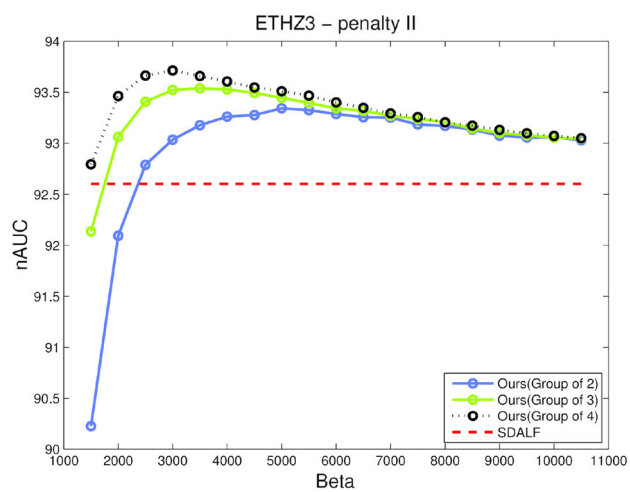
Similar results are shown in Figs. 7, 8, and 9 which contain experimental results of the SDC_{knn} method on the ETHZ

datasets. The improvements are analogous with roughly 0.7 improvement of nAUC for all the ETHZ1, ETHZ2, and ETHZ3 datasets. Accuracy enhancement on the ETHZ dataset is even better when applying our re-ranking method on the SDALF method. From Figs. 10, 11, 12 we can see more significant improvement when the nAUC is raised up to more than 1.0 for the ETHZ1, ETHZ2, and ETHZ3 dataset. Also similar to experiments in the VIPeR dataset, we can gain most nAUC enhancement with groups of four persons appearing concurrently.

The CUHK01 dataset is the dataset producing modest performance boost compared to the VIPeR and ETHZ dataset, with approximately 0.25 in nAUC growth (Fig. 13). The best group configuration is also different when groups of four give worst improvement and groups of three achieve the best.



(a) Penalty I



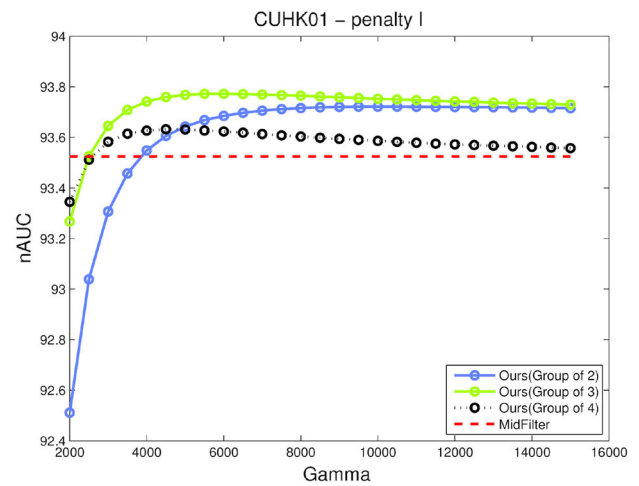
(b) Penalty II

Fig. 12 nAUC scores of the SDALF method and re-ranking method with different γ and β on ETHZ3

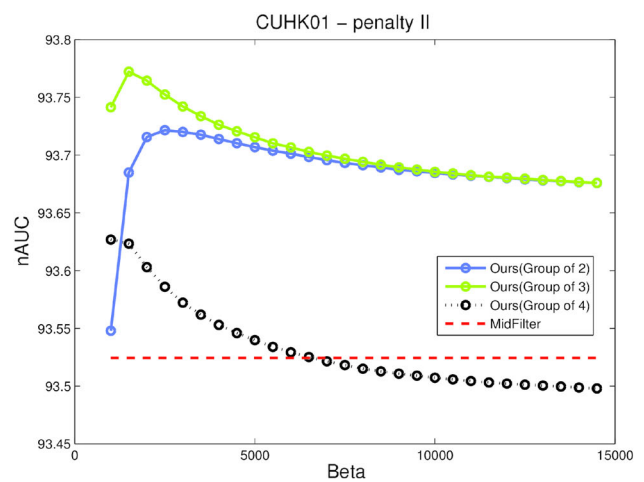
From the above results, we learn that very small γ and β cause a big drop in the results. This is because very small γ and β lead to big penalties which hurt the original score significantly. On the other hand, very large γ and β , which cause insignificant penalties, tend to make the performance converge to the original results.

In most of the cases, groups of four improved the performance the most. This can be explained by the fact that using a noisy list will badly affect other lists in the re-ranking procedure. By using 4 ranked lists at the same time, we have more information to balance the effect of noise from the lists.

In order to compare the effectiveness of the two penalty functions, Table 1 presents the most significant improvement of each configuration. There is no clear difference between the best performances of penalty I and penalty II. This means that even though the two penalties give different impacts on



(a) Penalty I



(b) Penalty II

Fig. 13 nAUC scores of the Mid-level filters method and re-ranking method with different γ and β on CUHK01

the final results (which can be seen through curves with different shapes in the figures), their improvement limits are similar.

5 Conclusion

In this paper, we proposed a re-ranking method which refines person re-identification results in the context of multiple people appearing concurrently in a camera. The experimental results with different state-of-the-art person re-identification methods on different datasets showed remarkable improvement when applying our method, especially when there are more people appearing at the same time. As a post-processing procedure, our proposed method can be applied to any state-of-the-art re-identification systems to boost

Table 1 Comparison between Penalty I and Penalty II

Method	Dataset	$k = 2$		$k = 3$		$k = 4$	
		Penalty I	Penalty II	Penalty I	Penalty II	Penalty I	Penalty II
SDALF	VIPeR	0.23	0.22	0.33	0.32	0.43	0.43
	ETHZ1	0.44	0.44	0.61	0.61	0.74	0.74
	ETHZ2	0.68	0.71	0.94	0.94	1.02	1.02
	ETHZ3	0.74	0.74	0.94	0.94	1.11	1.11
SDC _{knn}	VIPeR	0.21	0.21	0.34	0.33	0.46	0.46
	ETHZ1	0.26	0.26	0.55	0.55	0.69	0.69
	ETHZ2	0.31	0.30	0.53	0.51	0.70	0.70
	ETHZ3	0.48	0.48	0.61	0.61	0.73	0.73
QAF	VIPeR	0.18	0.20	0.05	0.06	0.20	0.20
MidFilter	CUHK01	0.20	0.20	0.25	0.25	0.11	0.10

The most significant improvement in each configuration is selected to show

their performance. For more accurate re-ranking, considering reliability of ranking lists would be a promising future study.

Acknowledgements This research is the output of the project Person re-identification using Semantic Features under Grant Number D2015-08 which belongs to University of Information Technology-Vietnam National University HoChiMinh City.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bedagkar-Gala, A., Shah, S.K.: A survey of approaches and trends in person re-identification. *Image Vis. Comput.* **32**(4), 270–286 (2014)
- Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: *BMVC*, p. 6 (2011)
- Chum, O., Philbin, J., Sivic, J., Isard, M., Zisserman, A.: Total recall: Automatic query expansion with a generative feature model for object retrieval. In: *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007*, pp. 1–8. IEEE (2007)
- Cui, J., Wen, F., Tang, X.: Real time google and live image search re-ranking. In: *Proceedings of the 16th ACM international conference on Multimedia*, pp. 729–732. ACM (2008)
- Ess, A., Leibe, B., Schindler, K., Van Gool, L.: A mobile vision system for robust multi-person tracking. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, pp. 1–8. IEEE (2008)
- Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2360–2367. IEEE (2010)
- Forsen, P.-E.: Maximally stable colour regions for recognition and matching. In: *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07*, pp. 1–8 (2007)
- Garcia, J., Martinel, N., Micheloni, C., Gardel, A.: Person re-identification ranking optimisation by discriminant context information analysis. In: *The IEEE International Conference on Computer Vision (ICCV)*, December 2015
- Gheissari, N., Sebastian, T.B., Hartley, R.: Person reidentification using spatiotemporal appearance. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol 2, pp. 1528–1535 (2006)
- Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: *IEEE International workshop on performance evaluation of tracking and surveillance*, Citeseer (2007)
- Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: *Proceedings of the 10th European Conference on Computer Vision: Part I. ECCV '08*, pp. 262–275. Springer, Berlin (2008)
- Jüngling, K., Bodensteiner, C., Arens, M.: Person re-identification in multi-camera networks. In: *2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 55–61. IEEE (2011)
- Layne, R., Hospedales, T., Gong, S. Person re-identification by attributes. In: *Proceedings of the British Machine Vision Conference*, pp. 24.1–24.11. BMVA Press (2012)
- Leng, Q., Ruimin, H., Liang, C., Wang, Y., Chen, J.: Person re-identification with content and context re-ranking. *Multimed. Tools Appl.* **74**(17), 6989–7014 (2015)
- Li, W., Zhao, R., Wang, X.: *Human Reidentification with Transferred Metric Learning*. Springer, Berlin (2013)
- Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: deep filter pairing neural network for person re-identification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 152–159 (2014)
- Li, Z., Chang, S., Liang, F., Huang, T., Cao, L., Smith, J.: Learning locally-adaptive decision functions for person verification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3610–3617 (2013)
- Liu, C., Loy, C.C., Gong, S., Wang, G.: Pop: person re-identification post-rank optimisation. In: *2013 IEEE International Conference on Computer Vision*, pp. 441–448 (2013)
- Liu, C., Gong, S., Loy, C.C., Lin, X.: Person re-identification: what features are important? In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) *Computer Vision ECCV 2012. Workshops and Demonstrations*, Volume 7583 of *Lecture Notes in Computer Science*, pp. 391–401. Springer, Berlin (2012)
- Liu, X., Song, M., Zhao, Q., Tao, D., Chen, C., Jiajun, B.: Attribute-restricted latent topic model for person re-identification. *Pattern Recognit.* **45**(12), 4204–4213 (2012)

21. Ma, B., Su, Y., Jurie, F.: Bicov: a novel image representation for person re-identification and face verification. In: Proceedings of the British Machine Vision Conference, pp. 57.1–57.11. BMVA Press (2012)
22. Ma, B., Su, Y., Jurie, F.: Local descriptors encoded by fisher vectors for person re-identification. In: Computer Vision–ECCV 2012. Workshops and Demonstrations, pp. 413–422. Springer (2012)
23. Nguyen, V.-H., Due Ngo, T., Nguyen, K.M.T.T., Duong, D.A., Nguyen, K., Le, D.-D.: Re-ranking for person re-identification. In: International Conference of Soft Computing and Pattern Recognition (SoCPaR), 2013, pp. 304–308. IEEE (2013)
24. Oreifej, O., Mehran, R., Shah, M.: Human identity recognition in aerial images. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 709–716 (2010)
25. Prosser, B., Zheng, W.-S., Gong, S., Xiang, T., Mary, Q.: Person re-identification by support vector ranking. In: BMVC, p. 5 (2010)
26. Schwartz, W.R., Davis, L.S.: Learning discriminative appearance-based models using partial least squares. In: Proceedings of the XXII Brazilian Symposium on Computer Graphics and Image Processing (2009)
27. Shen, X., Lin, Z., Brandt, J., Avidan, S., Wu, Y.: Object retrieval and localization with spatially-constrained similarity measure and $k - mn$ re-ranking. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3013–3020. IEEE (2012)
28. Tang, X., Liu, K., Cui, J., Wen, F., Wang, X.: Intentsearch: capturing user intention for one-click internet image search. IEEE Trans. Pattern Anal. Mach. Intell. **34**(7), 1342–1353 (2012)
29. Lindenbaum, M., Brand, Y., Avraham, T.: Transitive re-identification. In: Proceedings of the British Machine Vision Conference. BMVA Press (2013)
30. Zhao, R., Ouyang, W., Wang, X.: Unsupervised saliency learning for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3586–3593 (2013)
31. Zhao, R., Ouyang, W., Wang, X.: Unsupervised saliency learning for person re-identification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3586–3593 (2013)
32. Zhao, R., Ouyang, W., Wang, X.: Learning mid-level filters for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 144–151 (2014)
33. Zheng, L., Wang, S., Tian, L., He, F., Liu, Z., Tian, Q.: Query-adaptive late fusion for image search and person re-identification. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1741–1750 (2015)
34. Zheng, W.-S., Gong, S., Xiang, T.: Reidentification by relative distance comparison. IEEE Trans. Pattern Anal. Mach. Intell. **35**(3), 653–668 (2013)