Multimed Tools Appl (2019) 78:27067–27083
DOI 10.1007/s11042-017-4817-4

Local region partition for person re-identification

Huifang Chu[1] · Meibin Qi[1] · Hao Liu[1] · Jianguo Jiang[1]

Received: 23 February 2017 / Revised: 19 April 2017 / Accepted: 3 May 2017 /
Published online: 17 May 2017

Abstract Due to the different posture and view angle, the image will appear some objects that do not exist in another image of the same person captured by another camera. The region covered by new items adversely improved the difficulty of person re-identification. Therefore, we named these regions as Damaged Region (DR). To overcome the influence of DR, we propose a new way to extract feature based on the local region that divides both in the horizontal and vertical directions. Before splitting the image, we enlarge it with direction to increase the useful information, potentially reducing the impact of different viewing angles. Then each divided region is a separated part, and the results of the adjacent regions will be compared. As a result the region that gets a higher score is selected as the valid one, and which gets the lower score caused by pose variation and items occlusion will be invalid. Extensive experiments carried out on three person re-identification benchmarks, including VIPeR, PRID2011, CUHK01, clearly show the significant and consistent improvements over the state-of-the-art methods.

Keywords Person re-identification · Damaged region · Partition · Directional enlargement

---

✉ Meibin Qi
qimeibin@163.com

Huifang Chu
hfchu_hfut@163.com

Hao Liu
hfut.haoliu@gmail.com

Jianguo Jiang
jgjiang@hfut.edu.cn

[1] School of Computer and Information, Hefei University of Technology, Hefei, Anhui, 230009, China

1 Introduction

Person re-identification refers a task of associating the same person in cross camera views. But there are still many unsolved problems that due to changes in view angle, bodies pose, illumination, occlusion and background clutter. To address these difficulties, researchers are actively working on design feature [11, 15, 26, 42, 43, 63] and learning a distance [9, 12, 29, 34, 41, 44, 62], or both jointly [1, 2, 13, 14]. For feature representation, several effective approaches have been proposed, for example, Local Maximal Occurrence (LOMO) [13], GOG [26], KCCA [15], Mid-level [63], HOG3D [44], and SCNCD [43]. These hand-crafted features make impressive improvements in person re-identifications research. However, all of these feature descriptors are based on global appearance, or only taking the individual itself into account, without considering the attachment that captured by camera. Naturally, the interference caused by these areas that covered by the new appendages is ignored.

To solve these problems, we propose a novel framework for person re-identification by splitting the person picture in horizontal and vertical directions. This inspiration comes from SCSP [2], which divided target into four sub-regions in horizontal directions. The first difference is that we add the vertical direction of the regional division. It is very effective when a camera is shot in the front and another camera in the side. Moreover, the major improvement is that we do not regard the results of all regions as valid values. By contrast, we take a comparison in the neighboring region, the region getting a lower score will be treated as invalid one. This is effective for removing interference of Damaged Region.

Before splitting the target, we enlarge it with directional to add the useful information by bilinear interpolation algorithm. As a matter of fact, this measure combine with regional validity judgment can reduce the impact of viewpoint variation.

The features we have used are HSV and SILTP, and the extract method is the same as LOMO [13], the different is that we used the local mean instead local maximal. The mean value can increase the anti-interference of noise and improve the robustness, and reduce the randomness that brought by the maximum.

Our contributions can be summarized as follows:

- we present an efficient feature representation which combines HSV and SILTP after obtaining the mean value in the horizontal patches, and it is more robust than lomo [13].
- We proposed a novel framework by dividing the image into sub-region both in horizontal and vertical directions and used the higher score one as the valid region after comparison. This operation is more conducive to re-identifying work than not dividing the image or splitting the image in a single direction.
- We enlarge the image with direction before splitting them. The results show that the proposed model achieves an outstanding performance compared with the state-of-the-art person re-identification approaches.

The rest of this paper is organized as followed. Section 2, reviewing related works, and describing the framework of the proposed approach. Section 3 describe the details of our proposed method, include: how to extract the feature; the concept of Damaged Region (DR); the details of the image partition and directional enlargement. The experiments and results are in Section 4. We finally make a conclusion and discuss possible future works in Section 5.

2 Related work

Recently, Computer vision [19, 24, 38] has attracted wide attention,and image categorization, image cropping and segmentation and image recognition are important yet challenging tasks in this fields. Under the efforts of many researchers, many algorithms are proposed for image categorization [47, 49, 50, 55, 58, 60, 61]. More specifically, the [39, 40, 49, 55, 58] focus on the aerial image, and [47, 60, 61] belongs to fine-grained image categorization. In [50], Zhang proposed to learn object-shaped and directional receptive fields for image categorization. As one of the most basic photo manipulation processes, photo cropping [46, 56, 57] and image segmentation [48, 54, 59] attracted wide attention. In [17], the random forest model, which is useful for image segmentation and categorization, is constructed by semi-supervised node splitting. The image recognition mainly includes the action recognition [20, 21, 36, 64], activity recognition [16, 20, 21, 23, 25] and person re-identification. Semantic photo retargeting [51] and human fixations prediction [52] are the key technologies in action recognition. However, considering the limited storage space, we mainly discuss person re-identification in this paper.

Given a probe image, the person re-identification system aims at identifying a set of matching images from a gallery set, which is mostly captured by a different camera. It plays a crucial role in target tracking [4, 18, 27], target retrieval [6, 37] and predict work [22, 30]. In the past five years, person re-identification has attracted great attention and made great achievements with the joint efforts of the computer vision researchers. However, re-identifying results may be inaccurate and not robust due to appearance features variation caused by various viewpoint changes, occlusion, and individual movement factors. In addition, since people are mainly distinguishable by their clothing under surveillance setting, similar clothes among different persons add further challenges.

Recently, several research methods have been proposed for solving these problems in person re-identification. Design feature and distance metric are two key points. Many existing approaches try to build robust feature representations to describe a person's appearance under various conditions [11, 15, 26, 42, 43, 63]. Liao et al. [13] proposed an efficient feature representation called Local Maximal Occurrence (LOMO), using color and Scale Invariant Local Ternary Pattern (SILTP) histograms to represent picture appearance in a high dimension. This method locally constructs a histogram of pixel features, and then takes its maximum values within horizontal strips to overcome viewpoint variations while maintaining local discrimination. And mid-level filters (Mid-Level) in [63] was learned from patch clusters with coherent appearance gained by pruning hierarchical clustering trees to obtain view-invariant and discriminative features. In [20], it provided a mid-level feature representation for activities to captures the intrinsic properties for recognition work. In [26] a method called Gaussian of Gaussian (GOG) descriptor is proposed, which based on a shierarchical Gaussian distribution of pixel features. Using a complex algorithm in Riemannian mathematics, resulting in a considerable amount of time to extract the features. Chen et al. [2] partition an image into four non-overlap horizontal stripe regions, and each stripe region can be described by four visual cues which are organized as HSV1/HOG, HSV2/SILPT, LAB1/SILPT, LAB2/HOG. The optimization function adopt Alternating Direction Method of Multipliers (ADMM), which needs to spend more time to learn the coefficient matrix than the general metric learning methods. Besides, it is common to use multiple features [45, 53] to represent an object, and these features usually high-order correlated. Specifically, in the paper [45] a new multimodal feature integration framework was proposed, which construct a feature correlation hypergraph (FCH) to model the high-order relations among multimodal features.

Learning a robust distance metric is another important part in re-identification, which formulates the person re-identification as a distance metric learning problem where a transformation matrix is learned. And the distance is relatively small when extracted features represent the same person and big otherwise. In order to achieve this goal, many effective algorithms have been proposed these years. KISS Metric Learning(KISSME) [9], Cross-view Quadratic Discriminant Analysis (XQDA) [13], Metric Learning with Accelerated Proximal Gradient (MLAPG) [14],Top-push Distance Learning model (TDL) [44] are representative methods. These methods have raised recognition rate to a new level. Besides, Spatially Constrained Similarity function on Polynomial feature map(SCSP) [2] learns similiarity function with Mahalanobis distance and a bilinear similarity metric. Liao et al. [13] proposed a subspace and metric learning method called Cross-view Quadratic Discriminate Analysis (XQDA), which learned a discriminate low dimensional subspace by cross-view quadratic discriminate Analysis and got a QDA metric learned on the derived subspace at the same time. Li Zhang [62] presented to learn a discriminative null space for person re-identification, he used Null Foley-Sammon transform (NFST) to obtain the subspace by collapsing the training data points of each class to a single point. Our feature descriptor can reach high performance based on XQDA and NFST. We will compare our result with other feature descriptors on these two methods.

3 Our approach

For the two key issues of image-based person reidentification, we pay more attention to the first one, That is, the design of robust feature. We have designed a simple, but feasible method to get color and texture feature with high reliability and good computability. What's more, before extracting feature, We also have proposed to enlarge the image with directional, then we partition it into regions in a special way. In order to better explain the effectiveness of our method, we present the concept of Damaged Region(DR). The flowchart of our method is shown in Fig. 1. The details of each component are as follows.

3.1 Directional enlargement

As the resolution of the image from existing datasets is relatively low, therefore, we enlarge the image by bilinear interpolation algorithm to increase the useful information. What is

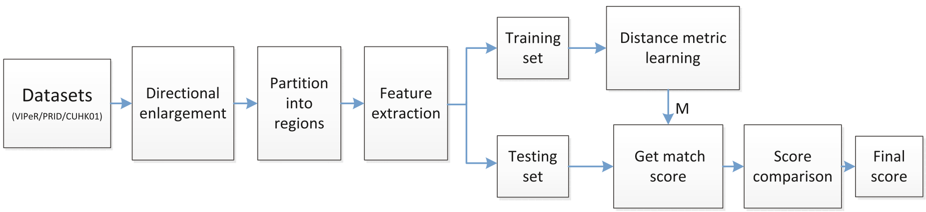

Fig. 1 Flowchart of the proposed method. We first enlarge the images with directional, and partition it into regions. Then we extract features for all images of the specific dataset. Lastly, the image will be split into two sets randomly. The training set is used for distance metric learning, and the testing set uses the measured matrix M which learned by the train set to calculate the match score for each region. The final score will be obtained after comparing the score in our own way

different with direct amplification is we adopt directional enlargement. The process of directional enlargement is described in detail in Fig. 2.

The yellow arrow represents the direction of the enlargement, and the red on behalf of the split direction. Before splitting the image in horizontal, we enlarge it in vertical direction with the size of 200 $*$ 80. When partitioning the image into vertical regions, the image is enlarged horizontally and scaled to 140 $*$ 120. Due to the need to take the score of global region into account, we normalize the image to 140 $*$ 80 in the case that the image doesn't need to be divided.

3.2 Damaged region

In realistic situation, pedestrians usually take something when they walk on the road, such as backpack, bag, book, etc. Due to the different poses and viewpoints, these attachments can't always be collected by different cameras. This means one of the images will appear some items that do not exist in another. Sometimes, some regions of targets will be covered by these items, which is harmful to identification work. Therefore, in this paper, we name these regions as Damaged Region (DR). For example, Fig. 3 shows some sample images that include Damaged Region from VIPeR dataset. The region marked with yellow wireframe is the Damaged Region.

With different viewpoints shown in Fig. 3a shot on the front and back, and Fig. 3b shot on the front and side, we can clearly find that in the second picture there are many new things that do not exist in the first. These new things covered pedestrian's own useful information, causing interference in matching with the first image. In Fig. 3c, under the influence of the angle and the posture, the dress of the person also emerges variation in local regions, these regions belong to Damaged Region as well.

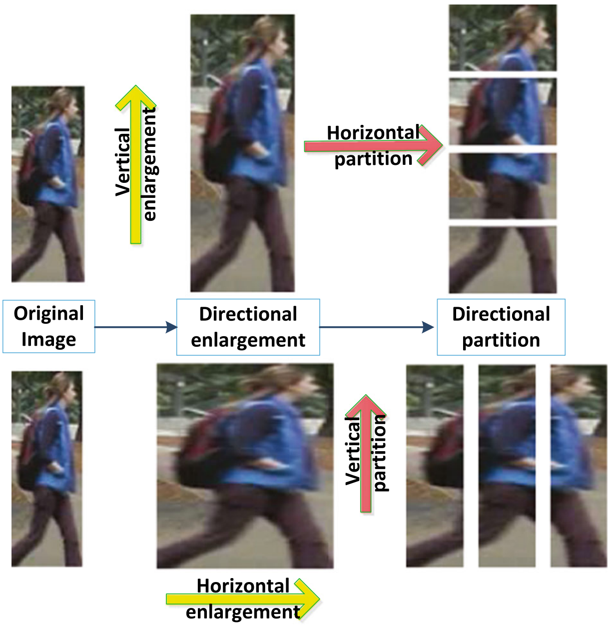

Fig. 2 The detail of directional enlargement. The *top* image is enlarged in vertical before partition into regions horizontally. And the same image below is enlarged in *horizontal* for splitting in *vertical*


(a) (b) (c)

Fig. 3 Examples of Damaged Region. The yellow wireframe in (**a**) (**b**) (**c**) is the DR. (**a**) Pictures are taken from the *front* and *back*. (**b**) Pictures are taken from the *front* and *side*. The DR in (**a**) and (**b**) is caused by attachments that captured in second camera. The DR in (**c**) is caused by the posture and the dress. Images in the same column represent the same person

There is a situation which we have ignored is that the pedestrian attachments appear in both pictures taken by different cameras with different locations, as shown in Fig. 4. Although these items also cover some areas that marked by green wireframe, because of its appearance in two images, these areas turn into the favorable factors for person re-identification, and it does not belong to the Damaged Region. We termed this region in Fig. 4 as Pseudo-damaged Region.

3.3 Partition into regions

After the introduction of the concept of the Damaged Region, we now discuss how to reduce the influence of it. In Fig 1, considering the damaged region is usually rectangular, we divide the picture into rectangular blocks so that the damaged area can be separated. The


(a) (b)

Fig. 4 Examples of attachments appear in two images. (**a**) Attachments cover the region in *horizontal*. (**b**) Attachments cover regions in *vertical*. Images in the same column represent the same person

R_{ri} represented the discrete region, where r is the number of regions and $i \in [1, r]$. The shape and size of the damaged area are also considered, so we have divided the images into horizontal four regions R_{41}, R_{42}, R_{43}, R_{44}, vertical 3 regions R_{31}, R_{32}, R_{33}. And through the experiment, the result is consistent with the analysis.

The Fig. 5 shows the example of how to eliminate the impact of damaged region by dividing the image in horizontal. R_{41}, R_{42}, R_{43}, and R_{44} are split in horizontal, the corresponding match scores are S_{41}, S_{42}, S_{43}, and S_{44}. We compare the S_{41} and the S_{42} to obtain the higher one, and S_{43} and S_{44} will be compared at the same time. As shown in Fig. 5, the higher scores are S_{41} and S_{43}, the homologous regions R_{41} and R_{43} are regarded as the valid region, The damaged region R_{42} usually gets a lower score, and is judged as an invalid one while compared with the others. The final score in horizontal when split into 4 regions S_4 is simply represented as:

$$S_4 = \begin{cases} S_{41} + S_{43} & S_{41} > S_{42}, S_{43} > S_{44} \\ S_{41} + S_{44} & S_{41} > S_{42}, S_{43} < S_{44} \\ S_{42} + S_{43} & S_{41} < S_{42}, S_{43} > S_{44} \\ S_{42} + S_{44} & S_{41} < S_{42}, S_{43} < S_{44} \end{cases} \tag{1}$$

Another important reason for our comparison between the R_{41} with R_{42}, R_{43} with R_{44} is: If there is no damaged region in the image, the first and fourth parts usually contain more background information, which is also a harmful factor in the person re-identification. Through regional validity judgment, the region that contains more background information also be regarded as an invalid region, and not should be added into the final result.

In the vertical direction, as shown in Fig. 3b, the damaged area is usually located on the left or right of the whole image area when it captured from frontal and profile view. In this case, the image is divided into three sub-regions R_{31}, R_{32}, R_{33}. We compare the S_{31} with S_{33}, then the higher one adds S_{32} as the final score, yielding vertical score S_3:

$$S_3 = \begin{cases} S_{31} + S_{32} & S_{31} > S_{33} \\ S_{33} + S_{32} & S_{31} < S_{33} \end{cases} \tag{2}$$

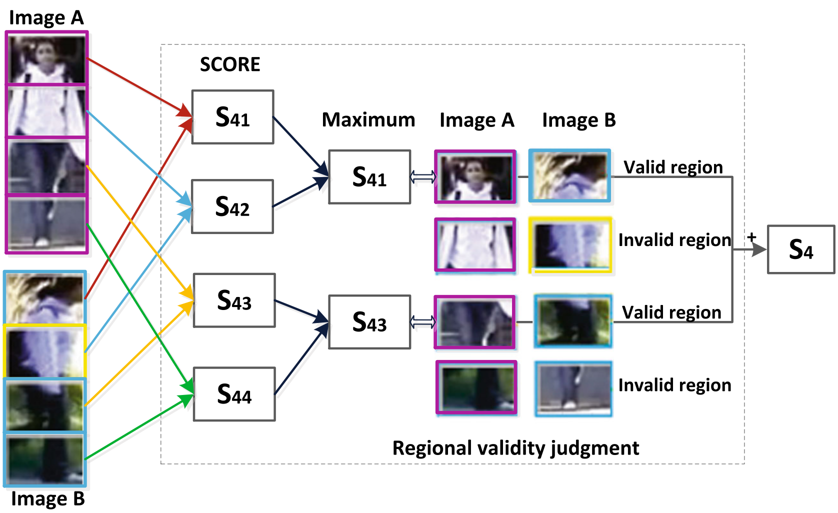

Fig. 5 Illustration of the regional validity judgment when the image is divided into regions in *horizontal*. The image **A** is in probe set and the image **B** is the same person that from gallery set

For Pseudo-damaged Region, Obviously, it is not feasible only to split the picture into four blocks in horizontal and three blocks in vertical. Consequently, we take the score of global image S_1 into account to compensate for the loss of Pseudo-damaged Regions that have been disabled. Furthermore, People's coat, whether its style or color, is usually more abundant. Generally, we judge people mostly from their coats, and the identification rate for upper body is more effective than those for lower body. So in order to make full use of the identification effect of the upper body, we will divide the person image into two regions in horizontal and compare the upper body and lower body to get the higher result S_2. From what has been discussed above, the overall similarity score S is given by:

$$S = S_1 + S_2 + S_3 + S_4 \tag{3}$$

3.4 Feature extraction

The color is an important feature for describing personal images. In this paper, we describe the object in HSV color space, which is insensitive to changes in light. Besides the color feature, we also apply the Scale Invariant Local Ternary Pattern (SILTP) descriptor for illumination invariant texture description, which is an improved LBP description operator. It is robust to region-wide noise, especially when the detection region is extremely dark or covered by shadows. The descriptor has a strong adaptability.

In the first place, an image is partitioned into many non-overlap horizontal and vertical stripe regions, each region serves as an independent individual. Then we divide each region into a collection of overlapped patches following [13], and extract HSV and SILTP histograms from each patch. We check all patches at the same horizontal location and take the average histogram of each feature among these patches. At last, the average histograms that belong to the same stripe region are connected. The resulting histogram achieves some invariance to viewpoint changes and robustness.

4 Experiments

4.1 Datasets and settings

Datasets Three widely used datasets are selected for experiments, including VIPeR, PRID and CUHK01.

VIPeR [5]**.** The VIPeR dataset is one of most challenging datasets for person re-identification task that has been widely used for benchmark evaluation. It contains two views of 632 persons. Each pair for a person is captured by different cameras with different pose, view angle, and illumination conditions. All images are scaled to $128 * 48$ pixels. We randomly select 316 persons for training, and the rest persons for testing. We repeat the procedure 10 times to get an average performance.

PRID2011 [7]**.** The PRID2011 dataset has two camera views as well. Camera A contains 385 persons, whilst camera B captures 749 persons. Only 200 persons appear in both views. We randomly select 100 persons to form the train set, while the remaining 100 persons of camera A to constitute the probe set and the remaining 649 of camera B are used as the gallery, likewise [62]. Experiments are repeated over the 10 times.

CUHK01 [10]**.** The CUHK01 dataset contains 971 persons from different camera view, and each person has two images in each camera view. Camera A captures the persons in frontal or back views while camera B captures them in the side view. All images were scaled to 160x60 pixels. We randomly divide the dataset under two settings single-shot and multi-shot. For each setting, the training/test split is repeated 10 times and the average of CMC result is reported.

Evaluation metrics. We use Cumulated Matching Characteristics (CMC) curve to evaluate the performance of person re-identification methods for selected datasets in this paper. this is an estimate of finding the correct match in the top n match, which is also known as rank-n. In order to more easily compare with published results, we report the cumulated matching result at selected rank-i ($i \in 1, 5, 10, 20$) in following tables.

4.2 Performance comparison

Our method can flexibly adapt any existing person re-identification methods of feature descriptors and metric learning models. In this section, we achieve our final results based on NFST. For fair comparison, we only compare with the results of using a single type of feature. The bold results in the table represent the best performances at present.

Results on VIPeR We first evaluate our method against the state-of-the-art on VIPeR. We are only comparing with the classic results in the past five years. Included KISSME [9], Mid-level [63], SCNCD [43], KLFDA [41], Sementic [32], CSL [31], XQDA [13], MLAPG [14], IDLA [1], S-CIR-CNN [35], SSDAL [33], NFST [62], S-SVM [62]. In addition, as paper [1, 33, 35] are the state-of-the-art methods based on deep learning, paper [9, 13, 14, 31, 32, 41, 43, 62, 62, 63] are the state-of-the-art methods based on handcrafted feature and metric learning. The best performance on VIPeR is S-SVM [62]. The comparison results are presented in Table 1. Our rank-1 matching rate 49.05% outperforms the most relevant one LOMO+NFST by 6.77% when a single type of features is used. Our method also improved the Rank-1 matching rate by 6.39% compared to the second best one S-SVM. The results show clearly our approach is superior to all deep learning methods on VIPeR dataset.

Results on PRID2011 We compare the state-of-the-art results [3, 8, 13, 15, 28, 41, 44, 62] reported on PRID2011 in Table 2. [3] belongs to the deep learning methods, and the

Table 1 Experimental evaluations on VIPeR dataset

Methods	Reference	Rank1	Rank5	Rank10	Rank20
KISSME	CVPR2012 [9]	19.60	–	62.20	77.00
Mid-level	CVPR2014 [63]	29.11	52.34	65.95	79.87
SCNCD	ECCV2014 [43]	37.80	68.05	81.20	90.04
KLFDA	ECCV2014 [41]	38.58	69.15	80.44	89.15
Sementic	CVPR2015 [32]	31.10	68.60	82.80	**94.90**
CSL	ICCV2015 [31]	34.80	68.70	82.30	91.80
XQDA	CVPR2015 [13]	40.00	68.13	80.51	91.08
MLAPG	ICCV2015 [14]	40.70	69.00	82.30	92.40
IDLA	CVPR2015 [1]	34.81	54.30	76.50	87.60
S-CIR	CVPR2016 [35]	35.76	67.00	83.00	–
SSDAL	ECCV2016 [33]	43.50	71.80	81.50	89.00
NFST	CVPR2016 [62]	42.28	71.46	82.94	92.06
S-SVM	CVPR2016 [62]	42.66	–	84.27	91.93
Ours	Proposed	**49.05**	**74.08**	**84.43**	93.10

Table 2 Experimental evaluations on PRID2011 dataset

Methods	Reference	Rank1	Rank5	Rank10	Rank20
RPLM	ECCV2012 [8]	15.00	32.00	42.00	54.00
KCCA	ACM2014 [15]	15.00	–	47.00	60.00
KLFDA	ECCV2014 [41]	22.40	46.50	58.10	68.60
Met-En	CVPR2015 [28]	17.90	39.00	50.00	62.00
XQDA	CVPR2015 [13]	26.70	49.90	61.90	73.80
MCP	CVPR2016 [3]	22.00	–	47.00	57.00
NFST	CVPR2016 [62]	29.80	52.90	66.00	76.50
TDL	CVPR2016 [44]	30.22	59.10	**74.04**	**88.43**
Ours	Proposed	**38.70**	**61.60**	71.80	81.10

rest are based on feature extraction and metric learning. TDL [44] get the best performance on PRID2011. Similar to the performance on VIPeR, the result is improved dramatically when compared with the previous state-of-the-arts. More specifically, we achieved 38.7% of the Rank-1 matching rate, whilst 30.22% for TDL, 29.8% for NFST, and 22.0% for MCP-CNN. Compared with the most related one NFST [62], our results increased by 8.9%, and 16% more than deep learning method MCP-CNN [3]. Comparing with TDL, We get higher results on rank-1 and rank-5, but not on rank-10 and rank-20. This is because TDL used multiple image frames for training and testing, while we just use only one of them for matching.

Results on CUHK01 Compared with VIPeR and PRID2011, CUHK01 are much bigger with thousands of training samples. We achieve the final result with single-shot and multi-shot, respectively. Mid-level [63], Semantic [32], IDLA [1], Met-En [28] are state-of-the-art methods with single-shot, while KLFDA [41], KCCA [15], MLAPG [14], NFST [62] are state-of-the-art methods with multi-shot. [13] achieve the result in two respects.

Single-shot For single-shot, 486 persons are randomly sampled for training and 485 for testing following [13, 26]. Results shown in Table 3 indicates that the proposed approach outperforms other feature descriptors such as Mid-level [63] and LOMO [13]. Our reports the best rank-1 recognition rate of 59.30%, with an improvement more than 10% over LOMO [13] which is regarded as the most relevant one, and an improvement of 11.77% over the deep learning method IDLA. It outperform the state-of-the-art method Met-En [28] with 5.9%, while both of them largely outperform the other existing state of the art methods.

Multi-shot For multi-shot, 485 persons are randomly selected for training and 486 for testing following [13, 26, 62]. Multi-shot matching scenario fused scores of multiple images of

Table 3 Experimental evaluations on CUHK01 dataset with single-shot

Methods	Reference	Rank1	Rank5	Rank10	Rank20
Mid-level	CCVPR2014 [63]	34.30	55.06	64.96	74.94
Semantic	CVPR2015 [32]	32.70	51.20	64.40	76.30
IDLA	CVPR2015 [1]	47.53	71.50	80.50	–
XQDA	CVPR2015 [13]	49.20	75.70	84.20	90.80
Met-En	CVPR2015 [28]	53.40	76.40	84.40	90.50
Ours	Proposed	**59.30**	**80.65**	**87.24**	**92.47**

Table 4 Experimental evaluations on CUHK01 dataset with multi-shot

Methods	**Reference**	**Rank1**	**Rank5**	**Rank10**	**Rank20**
KLFDA	ECCV2014 [41]	54.63	80.54	86.78	92.20
KCCA	ACM2014 [15]	56.30	80.66	87.94	93.00
XQDA	CVPR2015 [13]	63.21	83.89	90.04	94.16
MLAPG	ICCV2015 [14]	64.24	85.41	90.84	94.92
NFST	CVPR2016 [62]	64.98	84.96	89.92	94.36
Ours	Proposed	**70.45**	**87.92**	**92.67**	**96.34**

the same person by the sum rule. We compared available multi-shot results of KLFDA and KCCA from [62]. In Table 4, our approach outperforms NFST by achieving the recognition rate of 70.45%.

4.3 Effect of major components

We perform detail analysis of our approach on the VIPeR dataset. The two sets of Probe and Gallery each have 316 person images. The rank-1 indicates that the correct match result is at the top of all results. Therefore, only rank-1 matching rate which is the most important result is selected to compare in the following experiment.

Effect of new feature descriptor To better illustrate the effectiveness of our new feature descriptor, we compare ours with three classical feature descriptors, including LOMO, GOG, and SCSP. For fair comparison, we only choose one pixel feature vectors of $yM_\theta RGB$ for GOG descriptor. Similarly, we only use one visual cue that composed of color and texture features HSV/SILTP and divided into four stripes for the SCSP. At the same time, we select two methods to achieve the final result, one is NFST which belongs to subspace learning, and the other is XQDA that combine metric learning with subspace learning. The results of three features on XQDA come from the respective papers [2, 13, 26]. The results on NFST are achieved by our own according to the code provided. In Table 5, our feature descriptor evidently improves other feature descriptors both on XQDA and NFST.

Effect of partition regions In order to better illustrate the role of local region, we block the images both in the horizontal and vertical direction. The images are divided equally into regions R_r in both directions, and r= 2, 3, 4, 5, 6. It is necessary to investigate which partition is most effective. Our test is divided into two branches: considering the damaged region(C-DR) and not considering the damaged region (NC-DR). When the damaged area is not considered, we just sum the results of all blocks to get the final result. The Fig. 6 shows the result of Rank-1 matching rate. As shown in (a),(b), the R_2 and R_4 in horizontal

Table 5 Comparison with other feature descriptors

Methods	Reference	XQDA	NFST
LOMO	CVPR2015 [13]	40.00	42.28
SCSP	CVPR2016 [2]	37.09	36.08
GOG	CVPR2016 [26]	42.30	45.00
OURS	Proposed	**44.16**	**49.05**

are higher than others, and R_3 were significantly higher than those of the other blocks in vertical. What's more, as the number of r increasing, the recognition rate is gradually reduced both in horizontal and vertical.

Effect of global region We study the effect of pseudo-damaged regions by observing how the performance changes when the local region combined with global region in each direction. The experiments are also divided into two branches. From the results shown in Fig. 6(b),(c), we can obtain the following observations: (1) R_2 and R_4 in horizontal and R_3 in vertical still get the higher score. This further proves the rationality and effectiveness of our scheme. (2) Compared with (a),(b), we can find that (c),(d) results are significantly improved when taking the score of global region into a count. This is because the pseudo-damaged area is actually a favorable area, when testing the results of each separate partition like (a),(b), we usually treat the pseudo-damaged region as a damaged region, so that the result is damaged too.

Effect of horizontal-vertical collaboration Combining the partition in vertical and horizontal is a major innovation in the paper. It is necessary to verify the effectiveness of the combination, and how to combine to achieve the best results. Firstly, when testing the effect in horizontal, we keep the images divided into 3 regions in vertical, owing to R_3 get the highest result in (a),(b),(c),(d). As before, the R_2 regions and the R_4 are the best, and much higher than before in Fig. 6e,f. Therefore, when test in vertical, we maintain the picture divided into R_2 and R_4 in the horizontal direction, both block results are included in the final grade. The results get a much greater improvement over (c),(d). This fully illustrates the effectiveness of the combination of the two directions. In addition, it also explained those R_2 and R_4 in horizontal combine with R_3 in vertical is the best formation, which gets the best Rank-1 matching rate.

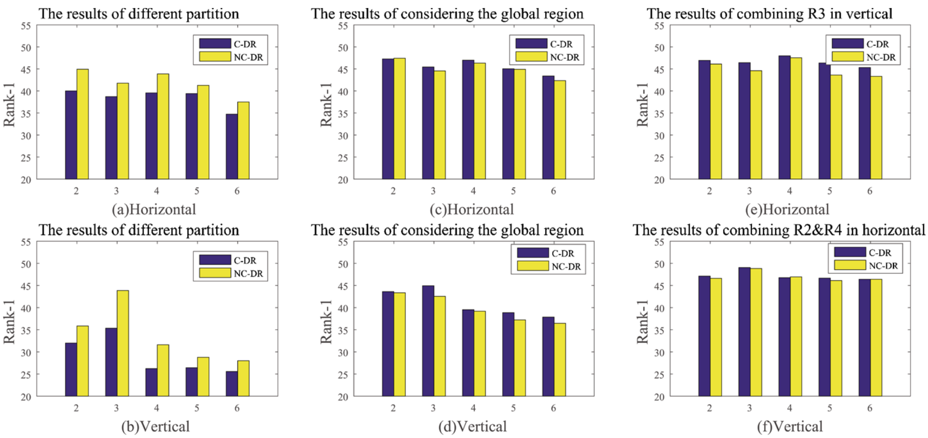

Fig. 6 Effect of major components. **1**.The abscissa represents the number of regions divided, and the ordinate is the Rank-1 matching rate. **2**. The *blue bar* means C-DR, and the *yellow bar* is NC-DR. **3**. (**a**),(**d**),(**e**) in the first row are the results of the horizontal partition, and (**b**),(**d**),(**f**) in second row are the results in vertical. **4**. (**a**), (**b**) show the results of partition regions in their respective direction. (**c**),(**d**) show the results when adding the global region. (**e**),(**f**) show the results of horizontal-vertical collaboration

Table 6 The results of image is enlarged

	NE	EH	EV	DE
Rank1	46.55	47.12	47.49	**49.05**

Effect of damaged region In this paper in order to prove the rationality of the damaged area, we carry out the experiments by divided them into two parts of C-DR and NC-DR. Through the previous experimental results in Fig. 6. We can draw the following conclusions:

(1) When the effect of each block is verified independently, (a),(b) clearly shows that the result of not considering the damaged region is higher than considered. In this case, compared to the direct added, once considered the damaged area, the role of entire region will be removed and much useful information will be reduced too. After all, only part of whole damaged area is covered, block can not be a complete removal of the specific part.
(2) When joining the global region, as shown in (c),(d),(e),(f), the results of C-DR is higher than NC-DR. Due to considering the damaged area and pseudo damaged area at the same time, the inadequacy caused by only considering damaged area will be made up, and improving the recognition rate. This also fully demonstrates the effectiveness of the damaged area and constraints.
(3) In all cases, the R_2 and R_4 in horizontal and R_3 in vertical always get the best results when compared with other block results respectively under the same experimental environment. This is consistent with the characteristics of the damaged region analyzed in Sec III.

Effect of directional enlargement In order to investigate the effect of amplification, we take experiment in four cases:

(1) Non-enlargement(NE)
(2) Enlarge it in horizontal(EH)
(3) Enlarge it in vertical(EV)
(4) Directional enlargement(DE)

The results as shown in Table 6, EH performs close to EV, EH and EV evidently improve NE by enlarging the image. DE takes advantages of the three, achieving 49.05% rank-1 matching rate.

5 Conclusions

In this paper, we have proposed a novel method by splitting the image into sub-regions in horizontal and vertical after enlarging it. The effectiveness of our method stems from the use of local region in two direction, which reduces the risk of mismatching, increases robustness to occlusion and is more flexible to handle pose variation. Our method also benefits from the global region that is complementary to the damaged regions. Extensive experiments on three benchmarks show that our method achieves the state-of-the-art performance. In the future, we will extend our approach by adopting more local region association strategies and by combining with other types of features and metric learning methods, which is expected to achieve better performance.

Acknowledgments This work was supported by the National Natural Science Foundation of China Grant 61371155.

References

1. Ahmed E, Jones M, Marks TK (2015) An improved deep learning architecture for person re-identification. In: Computer Vision and Pattern Recognition, pp 3908–3916
2. Chen D, Yuan Z, Chen B, Zheng N (2016) Similarity learning with spatial constraints for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1268–1277
3. Cheng D, Gong Y, Zhou S, Wang J, Zheng N (2016) Person re-identification by multi-channel parts-based cnn with improved triplet loss function. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1335–1344
4. Cui J, Liu Y, Xu Y, Zhao H (2013) Tracking generic human motion via fusion of low- and high-dimensional approaches. IEEE Trans Syst Man Cybern Syst 43:996–1002
5. Gray D, Brennan S, Tao H (2007) Evaluating appearance models for recognition, reacquisition, and tracking. IEEE International Workshop on PETS
6. Gupta P, Levy RC, Mattoo S, Remer LA, Munchak LA (2016) A surface reflectance scheme for retrieving aerosol optical depth over urban surfaces in modis dark target retrieval algorithm. Atmos Meas Tech 9(7):3293–3308
7. Hirzer M, Beleznai C, Roth PM, Bischof H (2011) Person re-identification by descriptive and discriminative classification. Lect Notes Comput Sci 6688(12):91–102
8. Hirzer M, Roth PM, Köstinger M, Bischof H (2012) Relaxed Pairwise Learned Metric for Person Re-identification. Springer, Berlin Heidelberg
9. Kostinger M, Hirzer M, Wohlhart P, Roth PM (2012) Large scale metric learning from equivalence constraints. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 2288–2295
10. Li W, Wang X (2013) Locally aligned feature transforms across views. Proc IEEE Conf Comput Vis Pattern Recognit 9(4):3594–3601
11. Li W, Zhao R, Xiao T, Wang X (2014) Deepreid: Deep filter pairing neural network for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 152–159
12. Li Z, Chang S, Liang F, Huang TS, Cao L, Smith JR (2013) Learning locally-adaptive decision functions for person verification. In: Computer Vision and Pattern Recognition, pp 3610–3617
13. Liao S, Hu Y, Zhu X, Li SZ (2015) Person re-identification by local maximal occurrence representation and metric learning. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 2197–2206
14. Liao S, Li SZ (2015) Efficient psd constrained asymmetric metric learning for person re-identification. In: IEEE International Conference on Computer Vision, pp 3685–3693
15. Lisanti G, Masi I, Bimbo AD (2014) Matching people across camera views using kernel canonical correlation analysis. In: Proceedings of the International Conference on Distributed Smart Cameras. ACM, pp 1–6
16. Liu L, Cheng L, Liu Y, Jia Y, Rosenblum DS (2016) Recognizing complex activities by a probabilistic interval-based model. AAAI Conference on Artificial Intelligence
17. Liu X, Song M, Tao D, Liu Z (2013) Semi-supervised node splitting for random forest construction. In: Proceedings of CVPR
18. Liu Y, Cui J, Zhao H, Zha H (2012) Fusion of low-and high-dimensional approaches by trackers sampling for generic human motion tracking. In: International Conference on Pattern Recognition, pp 898–901
19. Liu Y, Liang Y, Liu S, Rosenblum DS, Zheng Y (2016) Predicting urban water quality with ubiquitous data
20. Liu Y, Nie L, Han L, Zhang L, Rosenblum DS (2016) Action2activity: recognizing complex activities from sensor data. In: International Conference on Artificial Intelligence, pp 1617–1623
21. Liu Y, Nie L, Liu L, Rosenblum DS (2016) From action to activity: Sensor-based activity recognition. Neurocomputing 181:108–115
22. Liu Y, Zhang L, Nie L, Yan Y, Rosenblum DS (2016) Fortune teller: Predicting your career path. AAAI Conference on Artificial Intelligence
23. Liu Y, Zhang X, Cui J, Wu C (2010) Visual analysis of child-adult interactive behaviors in video sequences. In: International Conference on Virtual Systems and Multimedia, pp 26–33
24. Liu Y, Yu Z, Liang Y, Shuming L, Rosenblum DS (2016) Urban water quality prediction based on multi-task multi-view learning, IJCAI
25. Lu Y, Wei Y, Liu L, Zhong J, Sun L, Liu Y (2016) Towards unsupervised physical activity recognition using smartphone accelerometers. Multimedia Tools & Applications 1–19

26. Matsukawa T, Okabe T, Suzuki E, Sato Y (2016) Hierarchical gaussian descriptor for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1363–1372
27. Ning J, Yang J, Jiang S, Zhang L, Yang MH (2016) Object tracking via dual linear structured svm and explicit feature map. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 4266–4274
28. Paisitkriangkrai S, Shen C, Hengel AVD (2015) Learning to rank in person re-identification with metric ensembles. Comput Sci 1846–1855
29. Pedagadi S, Orwell J, Velastin S, Boghossian B (2013) Local fisher discriminant analysis for pedestrian re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 3318–3325
30. Shan Z, Xia Y, Hou P, He J (2016) Fusing incomplete multisensor heterogeneous data to estimate urban traffic. IEEE Multimedia 23:56–63
31. Shen Y, Lin W, Yan J, Xu M, Wu J, Wang J (2015) Person re-identification with correspondence structure learning. IEEE Int Conf Comput Vis. 3200–3208
32. Shi Z, Hospedales TM, Xiang T (2015) Transferring a semantic representation for person re-identification and search. In: Computer Vision and Pattern Recognition, pp 4184–4193
33. Su C, Zhang S, Xing J, Gao W, Tian Q (2016) Deep attributes driven multi-camera person re-identification. European Conference on Computer Vision
34. Tao D, Jin L, Wang Y, Yuan Y (2013) Person re-identification by regularized smoothing kiss metric learning. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1675–1685
35. Wang F, Zuo W, Lin L, Zhang D, Zhang L (2016) Joint learning of single-image and cross-image representations for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1288–1296
36. Wang W, Yan Y, Zhang L, Hong R, Sebe N (2016) Collaborative sparse coding for multiview action recognition. IEEE Multimedia Magazine 23(4):80–87
37. Xia Y, Chen J, Li J, Zhang Y (2016) Geometric discriminative features for aerial image retrieval in social media. Multimedia Systems 22:497–507
38. Xia Y, Nie L, Zhang L, Yang Y, Hong R, Li X (2016) Weakly supervised multilabel clustering and its applications in computer vision. IEEE Trans Cybern 46:1–13
39. Xia Y, Zhang L, Liu Z, Nie L, Li X (2016) Weakly-supervised multimodal kernel for categorizing aerial photographs. IEEE Trans Image Process
40. Xia Y, Zhang L, Tang S (2014) Large-scale aerial image categorization by multi-task discriminative topologies discovery. In: International Workshop on Internet-Scale Multimedia Management, pp 53–58
41. Xiong F, Gou M, Camps O, Sznaier M (2014) Person Re-Identification using Kernel-Based Metric Learning Methods. Springer International Publishing
42. Yan Y, Ni B, Song Z, Ma C, Yan Y, Yang X (2016) Person Re-identification via Recurrent Feature Aggregation. Springer International Publishing
43. Yang Y, Yang J, Yan J, Liao S, Yi D, Li SZ (2014) Salient color names for person re-identification. European Conference on Computer Vision
44. You J, Wu A, Li X, Zheng WS (2016) Top-push video-based person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1345–1353
45. Zhang L, Gao Y, Hong C, Feng Y, Zhu J, Cai D (2014) Feature correlation hypergraph: exploiting high-order potentials for multimodal recognition. IEEE T-CYB 44(8):1408
46. Zhang L, Gao Y, Ji R, Xia Y, Dai Q, Li X (2014) Actively learning human gaze shifting paths for semantics-aware photo cropping. IEEE T-IP 23(5):2235–2245
47. Zhang L, Gao Y, Xia Y, Dai Q, Li X (2015) A fine-grained image categorization system by cellet-encoded spatial pyramid modeling. IEEE Trans Ind Electron 62(1):564–571
48. Zhang L, Gao Y, Xia Y, Lu K (2014) Representative discovery of structure cues for weakly-supervised image segmentation. IEEE T-MM 16(2):470–479
49. Zhang L, Han Y, Yang Y, Song M, Yan S, Tian Q (2013) Discovering discriminative graphlets for aerial image categories recognition. IEEE T-IP 22(12):5071–5084
50. Zhang L, Hong R, Gao Y, Ji R (2016) Image categorization by learning a propagated graphlet path. IEEE T-NNLS 27(3):674–685
51. Zhang L, Li X, Nie L, Yan Y, Zimmermann R (2016) Semantic photo retargeting under noisy image labels. ACM TOMCCAP 12(3):37
52. Zhang L, Li X, Nie L, Yang Y (2016) Weakly supervised human fixations prediction. IEEE T-CYB 46(1):258
53. Zhang L, Song M, Bian W, Tao D, Liu X, Bu J, Chen C (2011) Feature relationships hypergraph for multimodal recognition. Springer, Berlin Heidelberg
54. Zhang L, Song M, Liu Z, Liu X, Bu J, Chen C (2013) Probabilistic graphlet cut: Exploiting spatial structure cue for weakly supervised image segmentation. In: Proceedings of CVPR
55. Zhang L, Song M, Sun L, Liu X (2012) Spatial graphlet matching kernel for recognizing aerial image categories. In: ICPR, pp 2813–2816

56. Zhang L, Song M, Yang Y, Zhao Q (2014) Weakly supervised photo cropping. IEEE Trans Multimedia 16(1):94–107
57. Zhang L, Song M, Zhao Q, Liu X, Bu J, Chen C (2013) Probabilistic graphlet transfer for photo cropping. IEEE T-IP 22(2):802–815
58. Zhang L, Wang M, Hong R, Yin BC, Li X (2016) Large-scale aerial image categorization using a multitask topological codebook. IEEE T-CYB 46(2):535–545
59. Zhang L, Yang Y, Gao Y, Yu Y, Wang C, Li X (2014) A probabilistic associative model for segmenting weakly-supervised images. IEEE T-IP 23(9):4150–4159
60. Zhang L, Yang Y, Wang M, Hong R, Nie L, Li X (2015) Detecting densely distributed graph patterns for fine-grained image categorization. IEEE T-IP 25(2):553–565
61. Zhang L, Yang Y, Zimmermann R (2015) Fine-grained image categorization by localizing tinyobject parts from unannotated images. In: ACM, pp 107–114
62. Zhang Y, Li B, Lu H, Irie A, Xiang R (2016) Sample-specific svm learning for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 1278–1287
63. Zhao R, Ouyang W, Wang X (2014) Learning mid-level filters for person re-identification. In: Computer Vision and Pattern Recognition, pp 144–151
64. Zhou Z, Song M, Zhang L, Tao D, Bu J, Chen C (2010) kpose: A new representation for action recognition. In: Asian Conference on Computer Vision, pp 436–447

Huifang Chu M. S. candidate at the Hefei University of Technology. She received her bachelor degree from Hefei University of Technology in 2015. Her research interests include video surveillance and person re-identification.

Meibin Qi a professor in the School of Computer and Information at Hefei University of Technology. He is received the BE degree in radio technology from Chongqing University in 1991, the ME and PhD degrees in signal and information processing from Hefei University of Technology in 2001 and 2007. His research interests include pattern recognition, video coding, video surveillance and the application of DSP technology.

Hao Liu received the M.Eng. degree in Communication and Information Systems from Hefei Univ ersity of Technology, where he is currently pursuing the Ph.D. degree. He is also a co-trained Ph.D. candidate with the National University of Singapore. His research interests include computer vision, image processing, pattern recognition, and deep learning. He is an reviewer of the ACM Transactions on Multimedia Computing, Communications, and Applications, the IEEE Transactions on Cybernetics and several other journals.

Jianguo Jiang a professor in the School of Computer and Information at Hefei University of Technology. He is received the BE degree in radio technology and ME degree in signal and information processing from Hefei University of Technology in 1982 and 1989. His research interests include digital image analysis and processing, distributed intelligent system and the application of DSP technology.