J. Vis. Commun. Image R. 29 (2015) 89–102

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

A person re-identification algorithm by exploiting region-based feature salience [☆,☆☆]

Yanbing Geng [a,c], Hai-Miao Hu [a,b,*], Guodong Zeng [a], Jin Zheng [a,b]

[a] Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
[b] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China
[c] College of Electronic and Computer Science and Technology, North University of China, Taiyuan, Shanxi 030051, China

ARTICLE INFO

Article history:
Received 29 August 2014
Accepted 4 February 2015
Available online 19 February 2015

Keywords:
Person re-identification
Region-based feature saliance
Salient color descriptor
Feature extraction
Feature fusion
Metric distance calculation
Illumination variation
Video surveillance

ABSTRACT

Due to the changes of the pose and illumination, the appearances of the person captured in surveillance may have obvious variation. Different parts of persons will possess different characteristics. Applying the same feature extraction and description to all parts without differentiating their characteristics will result in poor re-identification performances. Therefore, a person re-identification algorithm is proposed to fully exploit region-based feature salience. Firstly, each person is divided into the upper part and the lower part. Correspondingly, a part-based feature extraction algorithm is proposed to adopt different features for different parts. Moreover, the features of every part are separately represented to retain their salience. Secondly, in order to accurately represent the color feature, the salient color descriptor is proposed by considering the color diversity between current region and its surrounding regions. The experimental results demonstrate that the proposed algorithm can improve the accuracy of person re-identification compared with the state-of-the-art algorithms.

1. Introduction

Person re-identification is the task of establishing correspondences between observations of the same person in different videos. It faces many challenging issues like low resolution frames, time-varying light conditions, pose changes and partial occlusions under the uncontrolled complex environment. Conventional biometric traits such as face and gait are infeasible or unreliable owing to the scarce resolution of sensors. Usually, it is assumed that people among the different camera views can be effectively identified by some apparent information such as color, texture, edge and structure. The appearance-based person re-identification is widely used in the state-of-the-art algorithms.

Appearance-based person re-identification algorithm can be divided into two groups, namely the metric learning algorithm [20–29] and the feature representation algorithm [1–19,41–42]. Aiming to map the original feature to a new feature space, the first group learns a proper distance metric, which makes sure the feature distance between different instances for the same person is small in the new feature space. An optimal distance metric is learned through maximizing the inter-class variation while minimizing the intra-class variation in the literature [20,23–25], but it could easily cause over-fitted under the limited training samples. Zheng et al. [29] maximize the probability of a pair of true match having a smaller distance than that of a wrong match pair, which can alleviate the over-fitting problem. Xiong et al. [21] design classifiers to learn specialized metrics, which enforce features from the same individual to be closer than features from different individuals. However, enormous labeled training images should be obtained for the above mentioned methods, which may be hardly to implement in practice. Aiming at this deficiency, Liu et al. [22] propose a semi-supervised coupled dictionary learning, both labeled and unlabeled images are jointly learned in the training phase. However, these metric learning algorithms are inapplicable for practical surveillance applications, since the retraining must be carried out for per-dataset or per camera-pair.

The feature representation algorithms focus on seeking a distinctive and stable feature expression. Typical visual features are extracted for person re-identification, such as color, texture and shape. These features are always combined to improve the recognition rate. Feature extraction and multi-feature fusion are two

[☆] This paper has been recommended for acceptance by M.T. Sun.

[☆☆] This work was partially supported by the National Science Fund for Distinguished Young Scholars (No. 61125206), the National Natural Science Foundation of China (No. 61370121), the National Hi-Tech Research and Development Program (863 Program) of China (No. 2014AA015102), and Outstanding Tutors for doctoral dissertations of S&T project in Beijing (No. 20131000602).

[*] Corresponding author at: Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China.

E-mail address: frank0139@163.com (H.-M. Hu).

http://dx.doi.org/10.1016/j.jvcir.2015.02.001
1047-3203/© 2015 Elsevier Inc. All rights reserved.

main issues for the feature representation. We will analyze them in details.

Visual feature can be roughly categorized into global and local features. Color is the most commonly appearance feature for person re-identification. Global color features are encoded via chromatic histogram, since global color features often fail to characterize color distributions within the body part. Local color features are used to localize the color information through the patches segmentation. Mean-shift algorithm is used to segment an image into a series of patches in the literature [2], each patch has similar colors and HSV components are combined to represent the color information. However, this algorithm ignores the structure information of the person body, which will lead to the mismatch between torsos and legs. In order to deal with this problem, structure information is utilized as the spatial constraint by dividing a body into several asymmetrical parts (e.g., head, torso and legs) in the literatures [1,8–10]. According to [5], different parts are implicitly isolated and patches located in different parts are matched individually, which can efficiently reduce the mismatch between different parts. Asymmetry-based Histogram Plus Epitome (AHPE) feature is proposed to represent the global color histogram and local epitome information against the low resolution, occlusion and illumination variation in the literature [9]. However, the above algorithms often fail to differentiate persons with similar clothes and trousers. Inspired by human eyes of pedestrian identification relying on some salient regions, Zhao et al. [6] identify the matching pairs by means of the image salient patches. In their work, the salient patches are detected in an unsupervised manner and incorporated in patch matching to find reliable matches. But the salient patches may change due to illumination variation. Yang et al. [4] propose a salient color name based color descriptor (SCNCD) to analyze images by the semantic information, based on SCNCD, color distributions over sixteen color names in different color spaces are fused to address the illumination problem. Furthermore, because color is easily influenced by illumination variations across camera views, texture and shape are usually combined to model the human appearance. Schwartz and Davis [11] combine color, texture and edge features to represent the appearance, the partial least squares (PLS) is used for dimension reduction. But it fails to differentiate the salience among different features during the dimension reduction process.

Generally, the existing feature representation algorithms are insufficient in the following two aspects. Firstly, for the feature extraction, same features are adopted to describe different parts, on the assumption that these features are optimal for all parts. Since the person appearances captured by different cameras have obvious variations of posture and viewpoint, different features can be adopted for each part according to their respective characteristics. Furthermore, during the features fusion, there are various intrinsic meanings for each fused feature. Current multi-feature fusion algorithms fail to differentiate the salience among different features, which cannot make full use of the contribution of each feature.

Secondly, two kinds of local color descriptor, the stable patch color descriptor and the salient patch color descriptor are always used to describe the apparent color. However, both of them are less discriminative between the different appearances in the case of similar dress [1,6]. The salience should be considered not only for the local patch but also for the local color information. The detailed analysis of these two observations will be further discussed in Section 2.

Therefore, this paper proposes a person re-identification algorithm by fully exploiting the region-based feature salience (labeled as "RbFS"). In order to use the structural and spatial information, one human body is divided into the upper part and the lower part, and each part is further divided into multiple patches. The contributions of the proposed algorithm are two aspects. On the one hand, a part-based feature extraction algorithm is proposed to adopt different features for different parts according to the feature effectiveness and the part characteristics, so that the features for different parts are separately represented and each feature can retain its intrinsic meanings and salience. On the other hand, the salient color descriptor is proposed to embody the color representation and discrimination by detecting the salient color patch and considering the color diversity between current patch and its surrounding patches. The proposed algorithm is extensively evaluated on several public datasets, wherein images are captured in the video surveillance. The experimental results demonstrate that the proposed algorithm can significantly improves the accuracy of person re-identification compared with the-state-of-the-art algorithms.

The remainder of this paper is organized as follows. Two observations are elaborated in Section 2. The proposed algorithm is described in detail in Section 3. Section 4 summarizes the proposed algorithm. Section 5 evaluates the proposed algorithm on four datasets, including CAVIAR4REID [33], VIPeR [19], i-LIDS [39] and ETHZ [32]. And the paper is concluded in the final section.

2. Observations and justifications

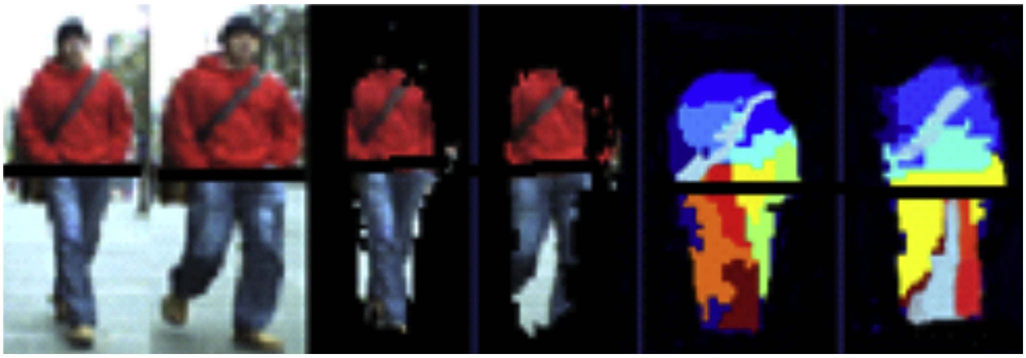
In this section, two observations are discussed in details and some experiments are carried out for justifications.

2.1. Feature extraction and fusion

The assumption of features universally optimal for the whole body is not desirable owing to the different characteristics of torso and legs. It is obvious from Fig. 1 that the lower region (e.g., legs) varies seriously whereas the upper region (e.g., torso) maintains relatively stable in general walking behavior.

Commonly, the appearance of human is usually characterized in three aspects, color, shape and texture. Color has proven to be effective for the task of person re-identification. It remains stable to the variations of posture and viewpoint even at lower resolutions. Texture and structure-shape information are complemented when color information degrades under illumination change. Each feature has its discriminative power, it is not powerful enough to characterize all apparent parts. Therefore the feature salience and effectiveness should be considered to represent different apparent part during the feature extraction. It is necessary for different regions to adopt different features. This conclusion can be proved through the following experiment on ETHZ and VIPeR datasets. For ETHZ, we randomly sampled 12 persons from each sequel. Six images for each person are selected, one for gallery set and the remainder for probe set. In addition, 100 persons are randomly chosen from VIPeR. For each person, one image forms gallery set and the other forms probe set.

In this experiment (as shown in Fig. 1), we apply human 2D vertical ellipse models [36] to partly remove the influence of the different background clutters, and then each body is divided into the upper part and the lower part by using adaptive body segmentation [5]. In order to make full use of the local detail information, each part is further segmented into several patches with Mean-shift [2]. Color feature (e.g., HSV value), texture features (e.g., Uniform Pattern Local Binary Patterns histogram, UPLBP [40]) and oriented gradient features (e.g., Histograms of Oriented Gradients, HOG [31]) are selected to describe each patch. Table 1 presents the rank 1 matching rate by using different feature on different parts. For the upper part, the average recognition ratios of adopting HSV value and HOG descriptor are 76% and 58% respectively, they are higher than those of adopting UPLBP. For the lower region, the



(a) Original images (b) Foreground object (c) Segmental patches

Fig. 1. Appearance discrepancy between the upper part and lower part of the same person.

Table 1
The rank 1 matching rate by using different feature on different regions.

Region	Dataset	HSV value (%)	HOG (%)	UPLBP (%)
Upper part	ETHZ SEQ1	87	71	63
	ETHZ SEQ2	75	65	42
	ETHZ SEQ3	91	78	67
	VIPeR	50	18	8
Avg.		***76***	***58***	*45*
Lower part	ETHZ SEQ1	78	50	69
	ETHZ SEQ2	60	46	55
	ETHZ SEQ3	88	53	75
	VIPeR	10	1	4
Avg.		***59***	*38*	***51***

HSV value and UPLBP are better than HOG. From Table 1, it can be observed that the color feature is simultaneously valid for both parts, but the edge structure feature (i.e., HOG) is more effective for the upper part, while the texture feature (i.e., UPLBP) is more suitable for the lower part. Therefore, the experimental results demonstrate that adopting different features for different parts can improve the accuracy of person re-identification.

Furthermore, in the multi-feature fusion, the salience of each fused feature should be made full use to improve the re-identification accuracy. However, the intrinsic meanings and sparsity of different feature descriptors are different [38]. Which can be observed from Fig. 2, wherein '*R*' represents one of the patches of legs. Fig. 2(a) presents the hue values (e.g., one component of HSV value) of '*R*', Fig. 2(c) presents the normalized LBP histogram of '*R*'. It is obvious that data of Fig. 2(c) is sparse compared with Fig. 2(a).

Eq. (1) is the conventional feature fusion method in [2,5], wherein the combined HSV value and UPLBP are directly inputted into Principal Components Analysis (PCA) for dimension reduction. However, this algorithm has deficiencies. PCA is utilized for dimension reduction and noise data removal only considering the data trait instead of the feature trait. As one of the combined feature, the UPLBP is possibly removed as noise due to its sparsity.

$$fs(R) = PCA\left(\bigcup_{k=1}^{m}\left(\bigcup_{i=1}^{3} fd_i(SR_k)\right)\right) \qquad R = \{upper, lower\},$$

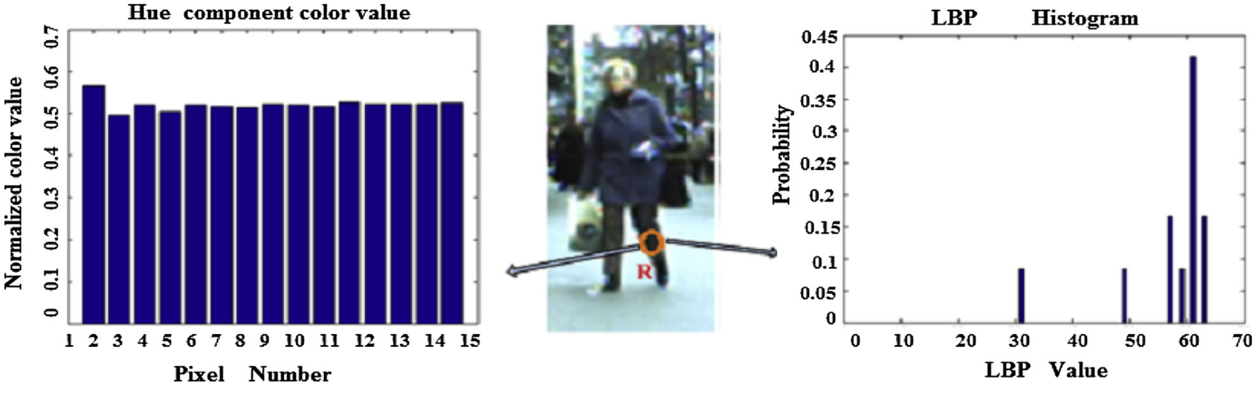
$$fd_i = \{HSV, UPLBP, HoG\} \tag{1}$$

where $fd(.)$ represents the selected single feature descriptor (e.g., HSV value, HOG or UPLBP), SR_k represents the kth patch R in the part, and $fs(.)$ represents the $PCA(.)$ processor on combined features directly.

Based on the above observation, we propose a part-based feature extraction algorithm, this algorithm takes the characteristics of different parts and the salience of different features into consideration. Specifically as follows: firstly, different features are extracted according to the characteristics of different parts, and then different features for each part are separately represented to make full use of the salience of different features during multi-feature fusion.

2.2. Color description

State-of-the-art approaches have segmented the full body into multiple patches for obtaining the local detail information. These patches are either assumed to be equally important or assigned



(a) HSV value of the region (b) Sample image (c) LBP histogram of the region

Fig. 2. Different feature descriptor of the patch.

global weights for matching all persons. It is inadvisable to assign global importance for all patches in many practical applications. As shown in Fig. 3, the red color patches play more important roles than gray patches in distinguishing one person from other person with a similar color dress. Motivated by the human visual attention, we believe several salient color patches play more important roles than others in describing an individual. Zhao et al. [6] share a similar view on this point. They propose an unsupervised salience learning to detect the salient patch, the patch salience of one image are computed by compared with neighbor patches located the corresponding coordinate of other images. But we think, for each image, human eyes can easily capture the salient patch of one image by observing the image itself, the patch will be salient once its feature obviously differs from those of its neighbor regions. The salient patch is detected by referring to its surrounding patches.

Moreover, conventional color descriptor is not representative and distinguishing. As shown in Fig. 4, it seems that red is the main color for the salient patch '*A*', while gray is the main color for its adjacent patch '*B*'. But patch '*A*' includes some other colors besides the red color. Hence, a representative color descriptor should be extracted to represent its color characteristics and distinguish from other patches.

Based on the above observations, it is necessary to detect the salient patch and extract the representative color descriptor for each salient patch. Therefore, a salient color descriptor is proposed to represent the local color information, which will be discussed in detail in the next section.

3. The proposed region-based feature salience exploration algorithm

This paper proposes a person re-identification algorithm by exploiting region-based feature salience, which has mainly two contributions based on two observations. As shown in Fig. 5, a part-based feature extraction is proposed to adopt different features for different part. On the other hand, the salient color descriptor is proposed to extract the representative color descriptor based on the salient patch detection.

3.1. The proposed part-based feature extraction

The proposed part-based feature extraction algorithm deems that features are extracted in accordance to the inherent appearance attributes. The upper parts have plenty of color information, and the color feature (e.g., HSV value) is an important feature for the upper parts. Shape feature (e.g., HOG) is formed through the calculation and statistics of local histogram of the oriented gradient. It is invariant to geometric and optical deformation in the subtle body movements. Since the upper parts are relatively stable and have slight posture changes during the normal walking, shape feature is preferable to describe the edge information. HSV value and HOG are suitable to describe the upper parts through the qualitative analysis.

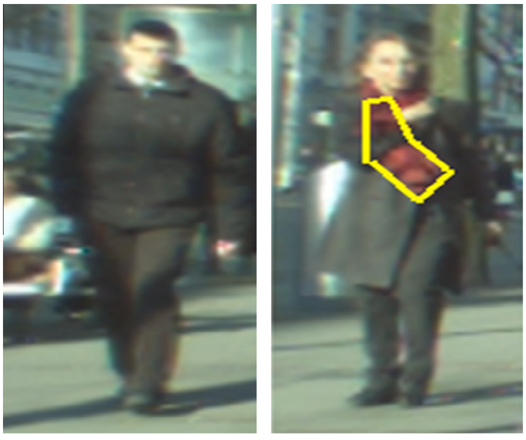


Fig. 3. Different persons with similar dresses.

For the lower parts, HSV value is also adopted to describe the color information. Since the lower parts of persons usually have obvious posture change, the edge of person may be significantly different among different images of the same person. HOG is a conventional feature to describe the oriented edge gradient. Thus, HOG fails to achieve a robust description of the lower parts.

On the other hand, since texture is a kind of local feature, texture of the lower parts is relatively stable among different images of the same person. Thus, texture can achieve a robust description of the lower parts. Moreover, Uniform Pattern LBP (labeled as "UPLBP") is adopted to characterize local texture. Compared with the original LBP, UPLBP can effectively avoid obtaining a sparse histogram by reducing binary patterns from 256 to 59. It is robust to illumination and rotation and it is relatively stable to posture changes. As shown in Fig. 6(b) and (c), where '*I1*' and '*I2*' belong to the same person, '*I3*' is the image of another person. UPLBP histograms of the same person are more approximate than that of the different person. Moreover, experiment on ETHZ SEQ3 is carried out to further demonstrate the validity of LBP in the lower body representation. In this experiment, UPLBP is extracted to represent each lower part. Euclidean distance is used for the distance metric. 'intra-person' refers to the lower part distance of the different images of the same person, 'inter-person' refers to the lower part distance of the different persons. Just as shown in Fig. 7. Distances of 'intra-person' are mostly lower than those of intra-person. This above analyses shows that HSV value and UPLBP are suitable to describe the lower parts.

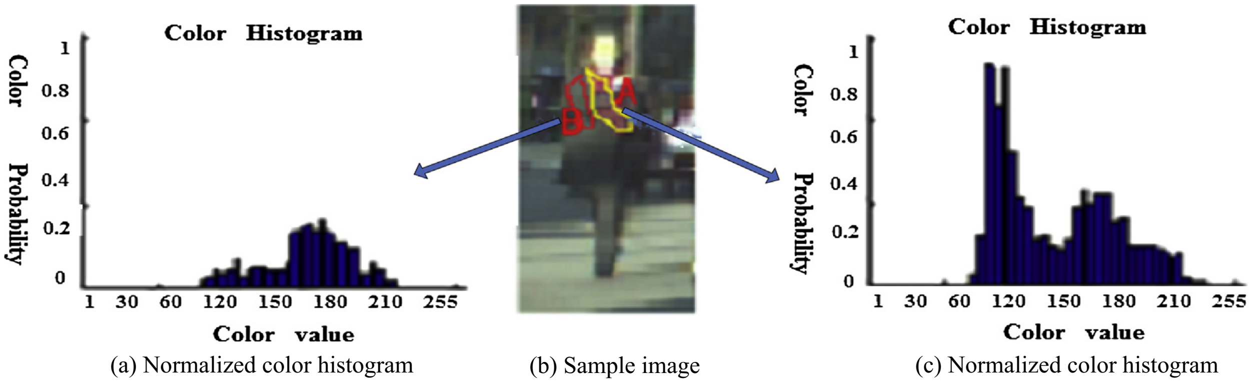
The conclusion is further demonstrated by the following experiment. Experiments are carried out on CAVIAR4REID, VIPeR, and ETHZ SQE3 datasets, because they present pose variety, occlusions and low resolution. We randomly select one image of each person for the gallery set, and the rest images for the probe set. The experimental settings are the same as the experiment in Section 2(a). For the upper part, rank 1 and 5 matching rate are shown in Table 2. It can be seen that the rank 1 average recognition ratios of adopting HSV value and HOG descriptor are 32.5% and 29% respectively, they are larger than 24.5% of adopting UPLBP, and the same conclusion can be drawn for the combined HSV and HoG. Therefore, HSV value and HOG are more suitable to describe of the upper part.

For the lower region, rank 1 and 5 matching rates are shown in Table 3. It is obvious that HSV value and UPLBP are better than HOG.

Therefore, for the feature extraction of the upper parts and the lower parts, the upper parts adopt both HSV value and HOG, while the lower parts adopt both HSV value and UPLBP in the proposed part-based feature extraction algorithm.

At the same time, in order to differentiate the salience among different features, *PCA* is firstly applied on each feature descriptor for dimension reduction, and then the new mapping features are combined. As shown in Eq. (2). '*R*' and 'SR_K' refer to the upper or lower part and its *m* patches, fd_i means *i*th fused feature.

$$fs_i(R) = \bigcup_{i=1}^{3}\left(PCA\left(\bigcup_{k=1}^{m} fd_i(SR_k)\right)\right), \quad R = \{upper, lower\},$$
$$fd_i = \{HSV, UPLBP, HoG\} \tag{2}$$



(a) Normalized color histogram (b) Sample image (c) Normalized color histogram

Fig. 4. The normalized color histogram of region '*A*' and '*B*'. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

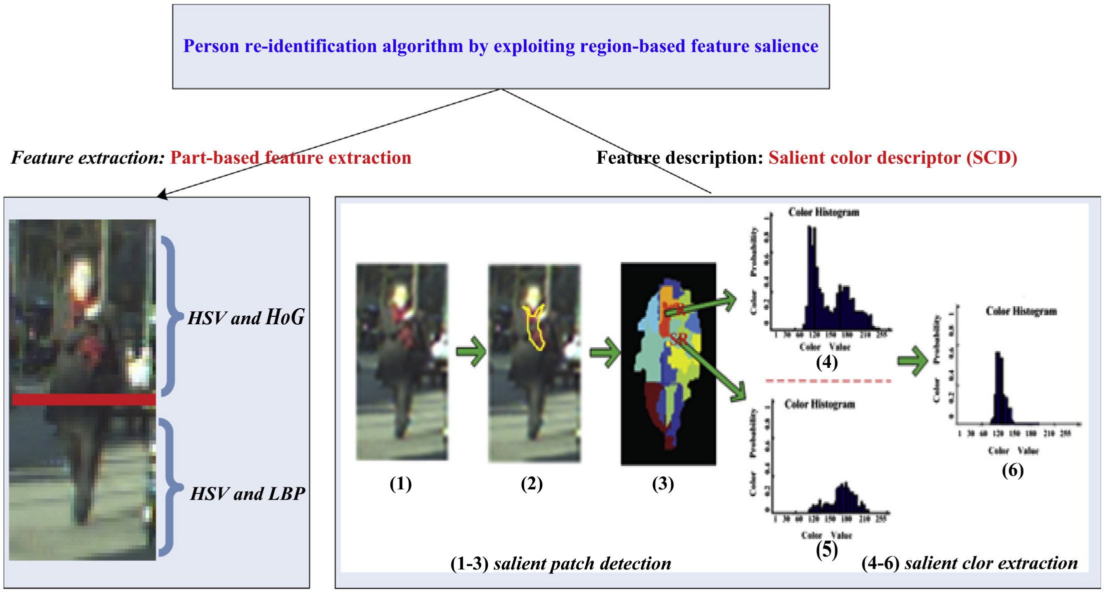

Person re-identification algorithm by exploiting region-based feature salience
Feature extraction: Part-based feature extraction
Feature description: Salient color descriptor (SCD)
HSV and HoG
HSV and LBP
(1)
(2)
(3)
R
SR
Color Histogram
Color Probability
Color Value
(4)
Color Histogram
Color Probability
Color Value
(5)
Color Histogram
Color Probability
Color Value
(6)
(1-3) salient patch detection
(4-6) salient clor extraction

Fig. 5. Diagram of the proposed algorithm. ((4) Color probability density of the salient patch '*R*'. (5) Color probability density of its neighbor patch '*SR*'. (6) Salient color probability density of '*R*'). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In this section, we propose a part-based feature extraction algorithm by taking the characteristics of different parts and the salience of different features into consideration. Liu et al. [34] has the similar the intention with our algorithm. They propose that some features may have a relatively strong discrimination, which can achieve a better description of a person. However, there are two major differences between the proposed algorithm and Liu et al. [34].

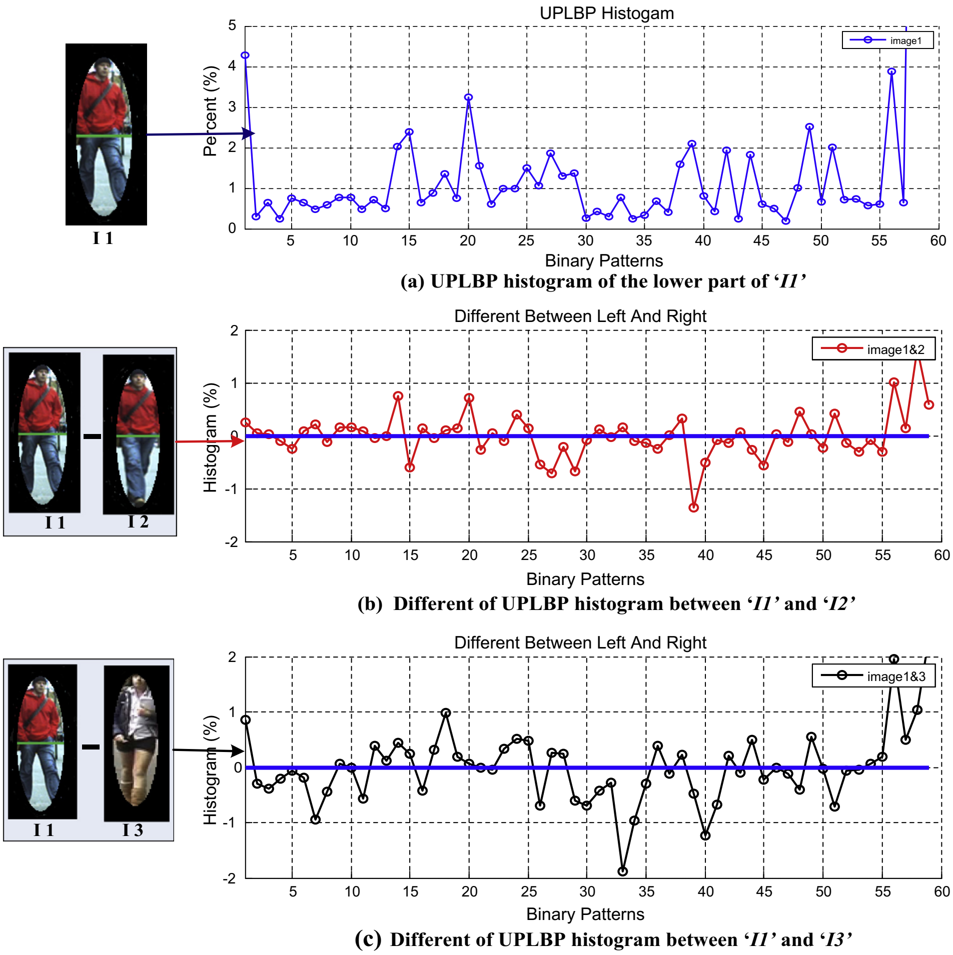
Firstly, Liu et al. [34] propose an unsupervised approach for learning adaptively feature weight. However, in practical application, the retraining must be carried out aimed at per-dataset or per camera-pair. We think different features can be adopted to each part (e.g., torso and legs) according to their respective characteristics, and no-training process is required. Secondly, our proposed feature fusion can retain the feature salience by separately representing each features of one region.

In the next section, the proposed salient color extraction algorithm will be introduced for the color discriminative representation of the sub-region.

3.2. The proposed salient color descriptor

The proposed salient color descriptor (SCD) aims at detecting the salient color patch and improving its weight in distinguishing a target from other individuals. Our salient color descriptor includes two main steps, namely the salient color patch detection and the representative color extraction for the salient patch.

The salient patch detection is separately carried out at each part (e.g., torso and leg) as a detect space constraint. Firstly, mean-shift is used to segment each part into several patches using color features. In contrast to [6] that image is densely segmented into a grid of local patches, our method guarantees the similar color for each patch. As shown in Fig. 8, n dimension matrix is obtained by calculating the weighted chromatic distances between any two patches. In our algorithm, the chromatic distance is obtained via Euclidean distance between two HSV histograms. The chromatic distance is weighted by one-dimensional Gaussian function $N(\mu, \sigma)$, where patches near the current detected patch count more, μ represents the centroid distance between two patches, and σ is confirmed by a priori value. The ith row of the matrix represents the metric distances between the ith patch and the others. Next, a threshold is set to decide whether other patches are similar with each detected patch or not, patch is removed by setting chromatic distance to zero when its chromatic distance is lower the threshold, since it is helpless and possibly interferential for the salient patch detection. Mean of the metric matrix is obtained as the threshold. Finally, the mean value is calculated for each row and formed n



(a) UPLBP histogram of the lower part of ‘*I1*’

(b) Different of UPLBP histogram between ‘*I1*’ and ‘*I2*’

(c) Different of UPLBP histogram between ‘*I1*’ and ‘*I3*’

Fig. 6. Difference of the UPLBP histogram of the lower parts.

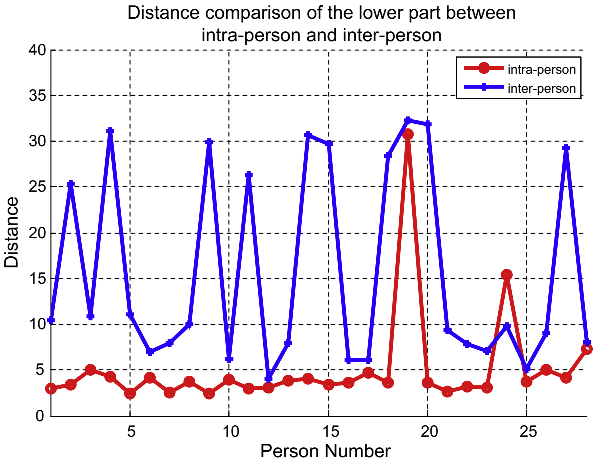


Fig. 7. Distance comparison of the lower part between intra-person and inter-person.

dimensions vector with descending order. Patches corresponding to the twenty percent of means are determined to be salient based on the experiments. The patch corresponding to lower deviation is relatively salient in the case of equal means.

To visualize the salience values of each part, Fig. 9 shows the effectiveness of salience detection on two dataset, red indicates larger weight. The proposed algorithm can detect the salient patches more efficiently.

In addition, the color discrepancy model $l(m)$ is defined to extract the representative color descriptor for each salient patch. As shown in Eq. (3), the HSV space is quantized into 8 bins for hue, 3 bins for saturation and 8 bins for value. $p_r(m)$ and $p_{sr}(m)$ refer to the 192 dimension normalized color histogram of the salient patch ‘*r*’ and its adjacent patches ‘*sr*’ respectively, m refers to the bin, *delta* is set to avoid the denominator for zero or the logarithm for zero. Originally, the mth color of salient patch is retained when $p_r(m) > p_{sr}(m)$. $l(m)$ is ranked in descending order and the top n $(n \leqslant m)$ colors are retained for representing the salient color of the salient patch. Colors corresponding to relatively small value $l(m)$ are removed owing to the possible interference of the salient patch.

Table 2
Rank *r* correct matches when using different feature on upper region.

Features		CAVIAR4REID	VIPeR ($p = 316$)	ETHZ SEQ3	Avg.
HSV value	$r = 1$	21.78	18.97	56.85	*32.50*
	$r = 5$	38.17	34.43	86.46	*53.02*
HoG	$r = 1$	23.45	10.80	53.93	*29.30*
	$r = 5$	40.39	28.50	78.09	*48.90*
UPLBP	$r = 1$	18.56	8.90	46.24	*24.50*
	$r = 5$	27.21	16.87	70.31	*38.13*
HSV and HoG	$r = 1$	25.13	20.78	57.85	*34.58*
	$r = 5$	42.90	40.84	84.82	*56.10*
HSV and UPLBP	$r = 1$	23.18	20.15	52.73	*32.02*
	$r = 5$	35.41	35.64	83.76	*51.60*

Table 3
Rank r correct matches when using different feature on lower region.

Features		CAVIAR4REID	VIPeR ($p = 316$)	ETHZ SEQ3	Avg.
HSV value	$r = 1$	16.51	6.8	38.15	*20.48*
	$r = 5$	31.39	10.2	57.31	*32.96*
HoG	$r = 1$	20.45	3.12	36.29	*19.9*
	$r = 5$	35.95	6.8	59.32	*34.02*
UPLBP	$r = 1$	26.56	2.3	38.57	*22.4*
	$r = 5$	41.34	5.5	71.43	*39.4*
HSV and HoG	$r = 1$	13.24	6.89	26.73	*15.62*
	$r = 5$	34.21	14.2	54.09	*34.16*
HSV and UPLBP	$r = 1$	18.33	5.32	47.57	*23.74*
	$r = 5$	36.56	12.7	65.2	*38.15*

$$l(m) = \log \frac{\max(p_r(m), delta)}{\max(p_{sr}(m), delta))}, \quad m = 1, 2, \ldots, 192 \tag{3}$$

In addition, we combine the HSV value and weighted hue histogram to represent the color information for each patch. The reasons are as follows: firstly, the HSV value means three component combination of HSV space for every pixel, which contains detail color information, but it is sensitive to the illumination and pose variation. Secondly, according to [37], hue is proven to be both lighting geometry invariant. Due to the uncertainty caused by small saturation in the computation of hue, hue is weighted by its saturation. The weighted hue histogram is complementary for the HSV value.

Some experiments are carried out to demonstrate the efficiency of the proposed SCD. We compare the proposed SCD with the original HSV value on the ETHZ dataset in the single-shot case. The experiment set is the same as that in Section 5. As shown in Fig. 10, experimental results demonstrate that the SCD algorithm has obvious improvement on the re-identification performance on ETHZ datasets.

4. The summary of the proposed algorithm

In this section, the proposed algorithm is summarized. Four main steps are image preprocessing, part-based feature extraction, salient color descriptor and the metric distance calculation.

4.1. Image preprocessing

Image is preprocessed to highlight the region of interest and to alleviate the influence of the background clutters. As shown in Fig. 11, all images are normalized to 79 pixels wide and scaled according to the original image. Next, Gaussian filtering is used for the effectively suppress noises. Since the background template cannot be effectively obtained, we apply the human 2D vertical ellipse model to partly remove the influence of the different background clutters. Finally the body is segmented into multiple patches by mean-shift.

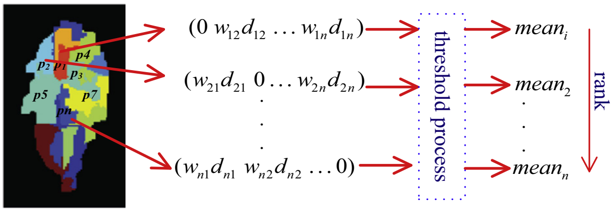

p4
p2
p1
p3
p5
p7
pn
$(0\ w_{12}d_{12}\ \ldots\ w_{1n}d_{1n})$
$(w_{21}d_{21}\ 0 \ldots\ w_{2n}d_{2n})$
$(w_{n1}d_{n1}\ w_{n2}d_{n2}\ \ldots 0)$
threshold process
$mean_i$
$mean_2$
$mean_n$
rank

Fig. 8. Flowchart of salient patch detection.

4.2. Part-based feature extraction

After the image preprocess, the whole body is divided into the upper part and lower part. Through qualitative analysis and quantitative verification, features are extracted according to the characteristics of each part. HSV value and HOG are extracted for the upper part. The lower part is represented by HSV value and UPLBP. PCA is applied on each feature descriptor to make full use of the salience of different features. The detailed calculation can be referred to Section 3.

4.3. Salient color descriptor

For each detected salient color patch, we retain the salient color information meanwhile removing the indistinctive color information. The Salient color is described by HSV values and 36 dimensions weighted Hue histogram according to [12]. The detailed calculation can be referred to Section 3.

4.4. The metric distance calculation

The metric distance of the proposed algorithm between the probe and gallery is calculated by summing the metric distance of upper part and lower part:

$$d_{rbfc}(p, g) = \alpha \cdot d_{rbfc}(p_{upper}, g_{upper}) + \beta \cdot d_{rbfc}(p_{lower}, g_{lower}) \tag{4}$$

where p and g represent the probe and gallery image, $d_{rbfc}(p, g)$ corresponds to the metric distance between two images, $d_{rbfc}(p_{upper}, g_{upper})$ and $d_{rbfc}(p_{lower}, g_{lower})$ are obtained via Earth Mover Distance (EMD [37]), α $(0 < \alpha < 1)$ is the weight of the upper part, and β $(0 < \beta < 1)$ is the weight for the lower part. Since apparent information for the upper part is richer than the lower part, in the experiment, α is fixed as 2/3 and β equals to 1/3.

Finally, our approach (i.e., RbFS) is combined with the existing approaches (i.e., MSCR [1] and SDC [6]). The matching of two signatures p and g is carried out by estimating the metric distance d.

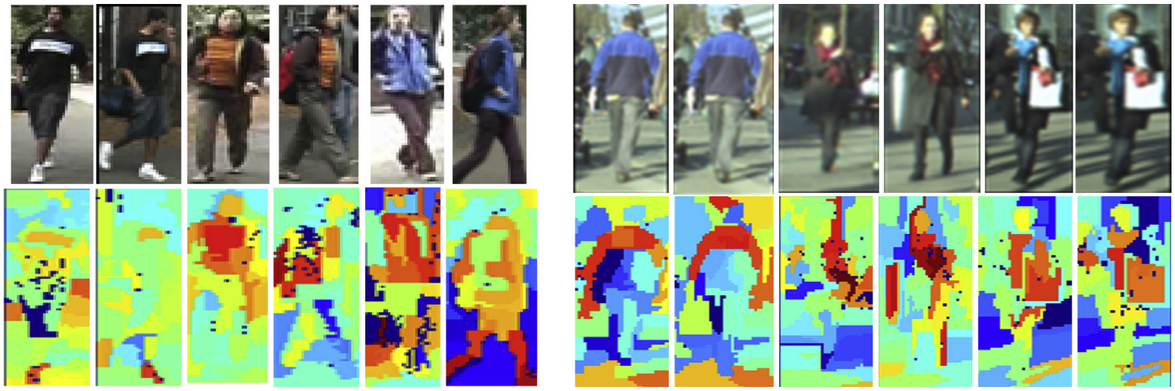
$$d(p, g) = \beta_{rbfc} \cdot d_{rbfc}(p, g) + \beta_{MSCR} d_{MSCR}(p, g) - \beta_{SDC} \cdot Sim_{SDC}(p, g) \tag{5}$$

where β_{rbfc} refers to the weight of region-based feature representation proposed by our algorithm, β_{MSCR} refers to the weight of MSCR [1] and β_{SDC} means the weight of SDC [6], d_{MSCR} corresponds to the metric distance proposed in [1], $Sim_{SDC}(p, g)$ corresponds to the similarity scores proposed in [6]. In this experiment, we fixed the values of the parameters as follows: β_{rbfc} equals to 1, β_{MSCR} equals to 0.7, and β_{SDC} equals to 1.

5. Experiments and discussion

In this section, the performance of the proposed algorithm is evaluated on four datasets, namely CAVIAR4REID, VIPeR, i-LIDS, and ETHZ. These datasets cover several challenging aspects of the person re-identification problem, such as pose variety, illumination changes, occlusions, image blurring and low resolution images. Fig. 12 shows some example images.

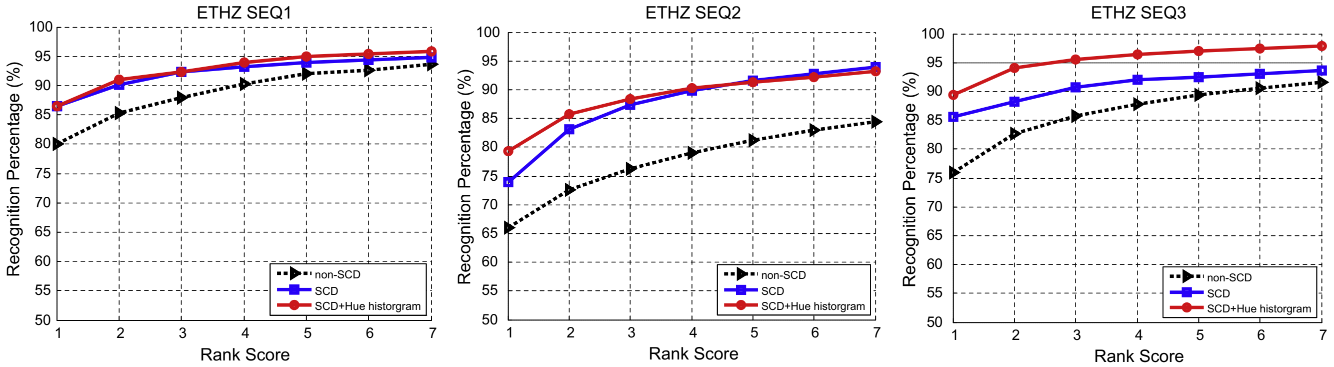
The performance is evaluated by the CMC curve and the Synthetic Recognition Rate (SRR) curve. The CMC curve represents the probability of finding the correct match in the top n candidates. The SRR curve depicts the probability that any of the m best matches is correct according to [1]. For the CMC curve, the proposed algorithm can be verified through two cases according to [8]. One is single-shot case (one image of each person is randomly selected and forms the gallery set, and the rest images form the probe set), the other is multiple-shot case (different images of the same person are randomly selected and form the gallery set, the rest images form the probe set.). The single-shot case is labeled as



(a) VIPeR

(b) ETHZ

Fig. 9. Qualitative validation of our salience detection (red indicates large weights). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

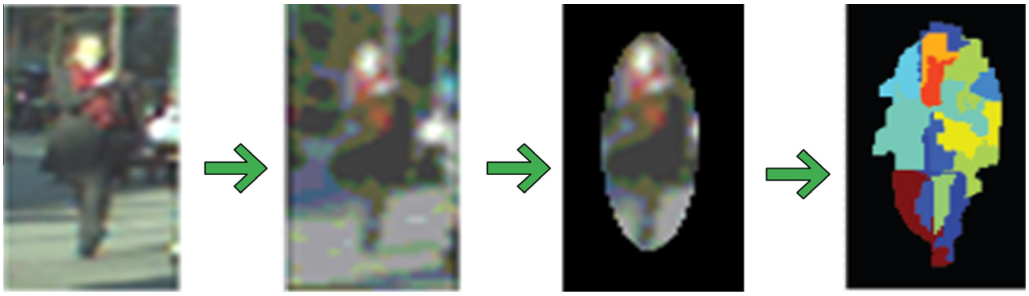


(a) CMC on ETHZ SEQ1

(b) CMC on ETHZ SEQ2

(c) CMC on ETHZ SEQ3

Fig. 10. Performance comparison for different color descriptors on ETHZ dataset in single-shot case.



(a)

(b)

(c)

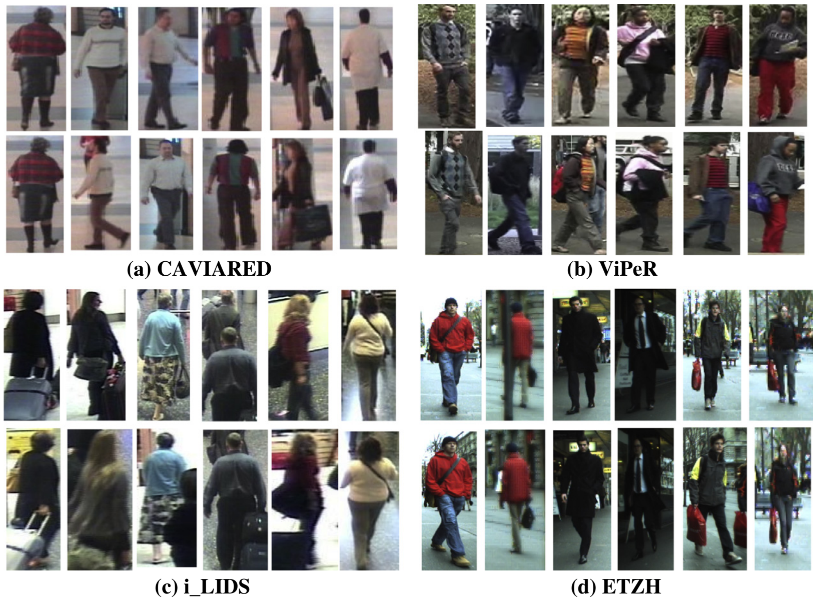
(d)

Fig. 11. mage preprocessing. ((a) Original image, (b) Image after Gaussian filtering of (a). (c) Image after removing part of background interference by human 2D vertical ellipse. (d) Segmentation result of (c)).

the SvsS case and the multiple-shot case is labeled as the MvsS case in this paper. For the SvsS case, the average performance over 10 trials is used to estimate the re-identification accuracy. The MvsS case is carried out by considering $N = 2, 3, 5$, and the re-identification performance is calculated by the average result of 100 independent trials. Figs. 13–21 show the results in the SvsS and MvsS case on different datasets.

The proposed algorithm is firstly evaluated on CAVIAR4REID dataset, which contains 1221 images of 72 pedestrians extracted from different cameras in the shopping center. Besides the illumination changes, pose variation, and viewpoint difference, the primary challenge is the low resolution with image sizes ranging from 17 × 39 to 72 × 144. Fig. 13 shows the CMC curve and SRR curve obtained by the proposed algorithm. For the CMC curve, the rank 1 correct matches is 32.4% in the SvsS case, and up to 56.7% in the MvsS case ($N = 5$). This means that the performance can be increased by adding to more instances to signature. The same conclusion can be drawn by the SRR curve.

In Fig. 14(a), the proposed algorithm is compared with SDALF [1] and eSDC_knn [6] in the SvsS case ($N = 1$). Since there is no experimental results on this dataset for SDALF and eSDC_knn, the results are calculated according to their source codes. The rank 1 matching rate of the proposed algorithm can achieve 32.70%, it has an obviously improvement compared with SDALF, and it is

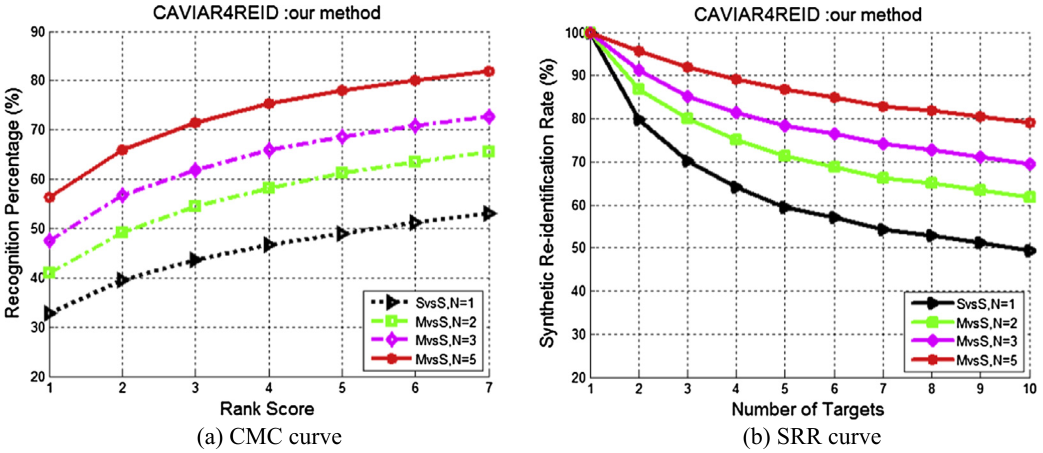

(a) CAVIARED

(b) ViPeR

(c) i_LIDS

(d) ETZH

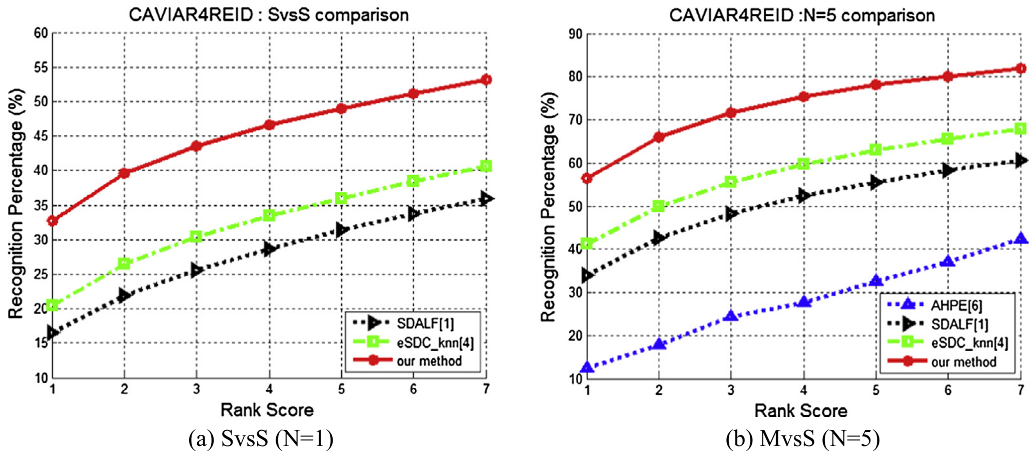
Fig. 12. Example images of different datasets.



(a) CMC curve

(b) SRR curve

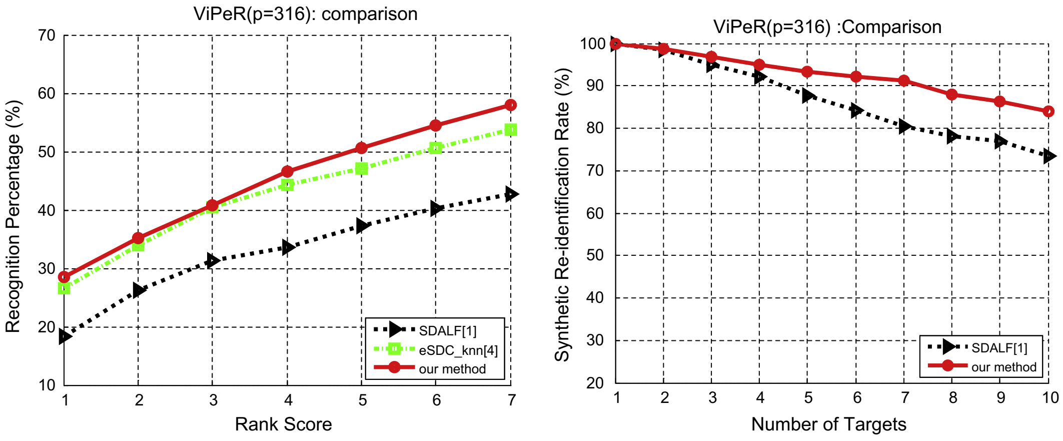
Fig. 13. Performances comparison of the proposed algorithm on CAVIAR4REID dataset.



(a) SvsS (N=1)

(b) MvsS (N=5)

Fig. 14. Performances comparison on CAVIAR4REID dataset.



(a) CMC curve (b) SRR curve

Fig. 15. Performances comparison on VIPeR dataset.

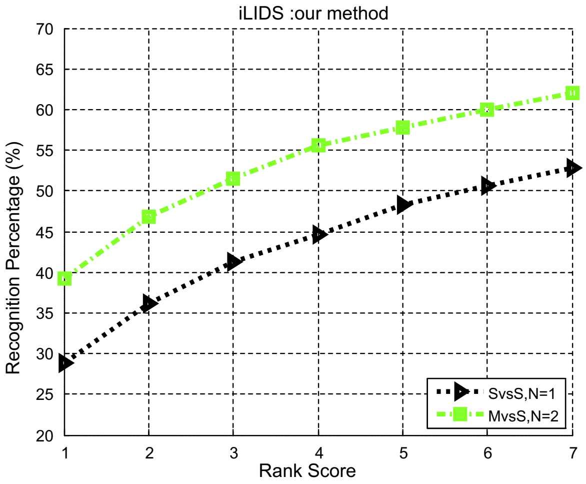
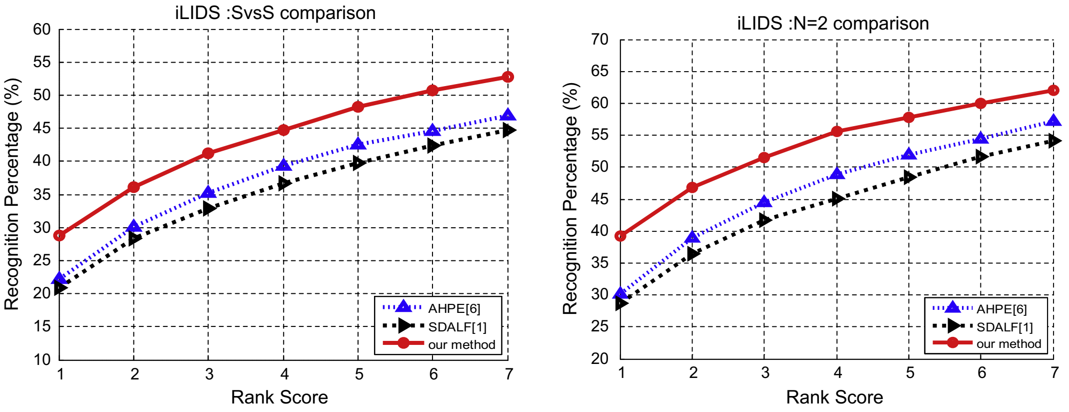


Fig. 16. Performances comparison on i_LIDS dataset (SvsS case ($N = 1$) and MvsS case ($N = 2$)).

improved by 12.25% in comparison to eSDC_knn. Similar conclusion is drawn in Fig. 14(b). The proposed algorithm is compared with SDALF, eSDC_knn and AHPE [9] in the MvsS case ($N = 5$). The experiment results demonstrate that for low resolution images, the proposed algorithm can improve the recognition performance compared with the state-of-the-art algorithms.

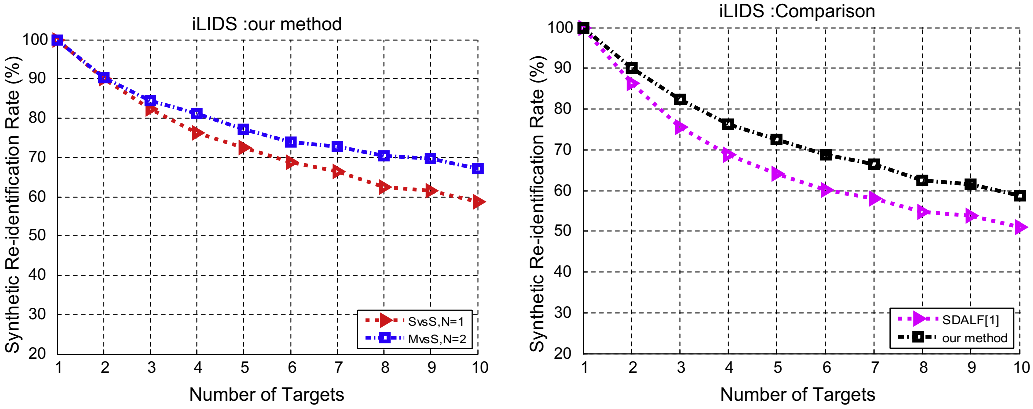
The performance is further evaluated on VIPeR dataset. It contains 1264 images of 632 person pairs captured from two cameras in outdoor scenes. It is the most challenging dataset for person re-identification owing to the dramatically viewpoint changes, obvious pose variations, and illumination differences. Since there are only two examples for each pedestrian, the experiment is carried out in the single-shot cases. The CMC and SRR curves are depicted by comparing our algorithm with the SDALF and eSDC_knn in Fig. 15. Following the experimental methods mentioned by SDALF and eSDC_knn, the average result of 10 repeated trials is obtained to evaluate the proposed algorithm, and 316 person pairs are randomly selected for each trial. Rank 1 matching rate is improved by 8% than SDALF, and the matching rate at rank twenty-six is around 80%, and there is a slightly improvement compared with eSDC_knn.

The i-LIDS dataset is captured at an airport arrival hall under multi-cameras conditions in a real scene. It contains 476 images of 119 persons. In addition to the obvious illumination changes, the main characteristics are the serious occlusion and incomplete appearance. In contrast to CAVIARED dataset, it presents more occlusion and incomplete appearance examples than CAVIARED. As shown in Fig. 16, the rank 1 correct matches is 28.29% in the SvsS case, and up to 39.96% in the MvsS case ($N = 2$).



(a) SvsS (N=1) (b) MvsS (N=2)

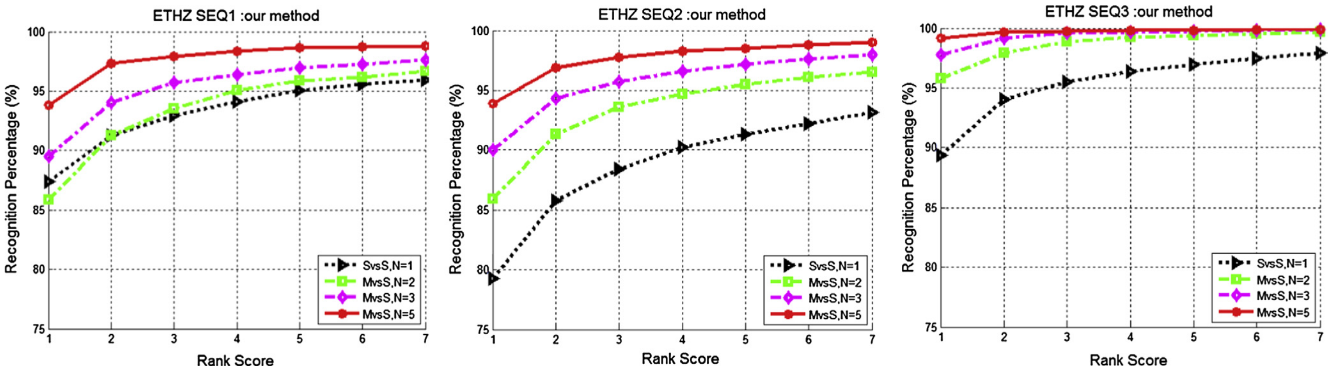
Fig. 17. CMC curve on i_LIDS dataset.



(a) SvsS (N=1)

(b) MvsS (N=2)

Fig. 18. SRR curve on i_LIDS dataset.

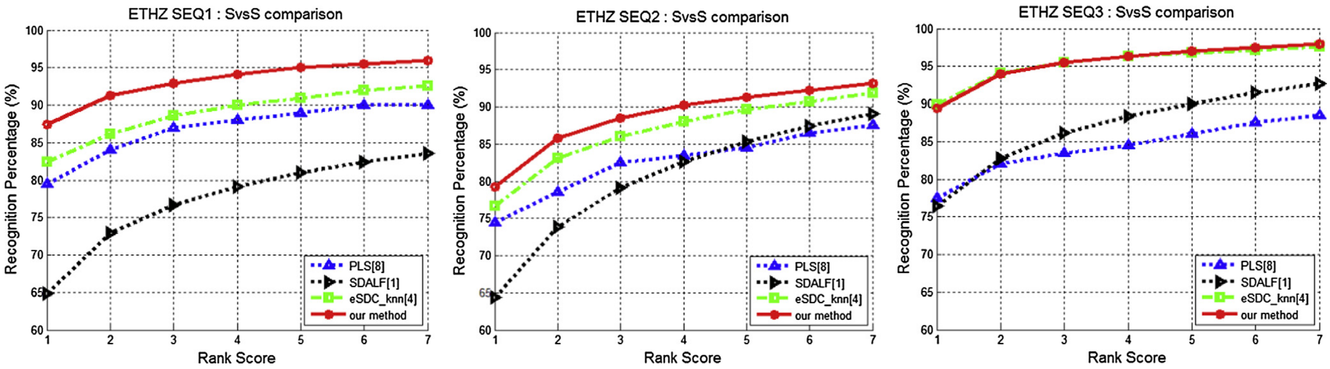


(a) CMC on ETHZ SEQ1

(b) CMC on ETHZ SEQ2

(c) CMC on ETHZ SEQ3

Fig. 19. Performances comparison in the SvsS ($N = 1$) case and MvsS ($N = 2, 3, 5$) cases.



(a) CMC on ETHZ SEQ1

(b) CMC on ETHZ SEQ2

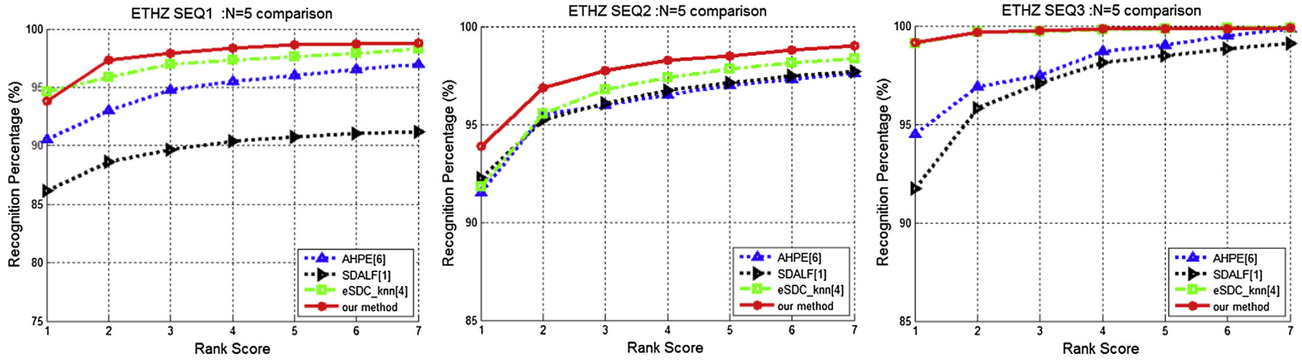
(c) CMC on ETHZ SEQ3

Fig. 20. Performances comparison on ETHZ dataset.

The proposed algorithm is compared with AHPE and SDALF in Fig. 17. AHPE and SDALF are proved to be relatively robust for occlusions in their literatures. The rank 1 match ratio is improved approximately by 8% than SDALF in the SvsS and MvsS case. This means that the proposed algorithm is more effective to deal with occlusions and incomplete appearance owing to the feature extraction for each part separately.

The re-identification performance on ETHZ is shown In Fig. 19. The ETHZ dataset contains three image series: ETHZ SEQ1 contains 4857 images of 83 persons, ETHZ SEQ2 contains 1936 images of 35 persons, and ETHZ SEQ3 contains 1762 images of 28 persons. The image sizes vary 30 ∗ 80 and 181 ∗ 402 pixels, and different images of one person have obvious posture changes and serious occlusions. We do not obtain the best results in the SvsS case for ETHZ SEQ2, but 93% rank 1 matching rate can be obtained in the MvsS case. A similar trend can be observed for ETHZ SEQ2 and ETHZ SEQ3.

Fig. 20 shows the performance comparison with SDALF, eSDC_knn and PLS [11] in the SvsS case ($N = 1$) on ETHZ datasets. For ETHZ SEQ1, the rank 1 matches ratio can achieve 86.7%, it is



(a) CMC on ETHZ SEQ1 (b) CMC on ETHZ SEQ2 (c) CMC on ETHZ SEQ3

Fig. 21. Performances comparison in the MvsS ($N = 5$) case on ETHZ dataset.

improved by 3% comparing with eSDC_knn, and there have obvious improvement compared with SDALF. For ETHZ SEQ2, since most images are low resolution, the rank 1 correct matches ratio is also improved by 2.5% comparing with eSDC_knn. Because ETHZ SEQ3 dataset is less challenging, we can obtain an approximate result with eSDC_knn. As shown In Fig. 21, the similar conclusion can be drawn by further comparing with SDALF [1], eSDC_knn [6], and AHPE [9] in the MvsS case ($N = 5$), although AHPE is proposed to solve the low resolution problem, our algorithm is obvious better than SDALF and AHPE.

In addition, to validate the matching performance of only the proposed algorithm (RbFS), we analyze the top r rank matching rate on the following datasets. As shown in Table 4, we compare the results with SDALF. SDALF includes three complementary features, their proposed RHSP feature and two existing features (i.e., wHSV and MSCR). Our proposed algorithm obviously improves the matching rate on the three datasets.

Our proposed algorithm outperforms eSDC_knn on CAVIAR4-REID. eSDC_knn is formed by combining their proposed SDC feature and two existing features(i.e., wHSV and MSCR). There is a slight decline compared with eSDC_knn on the VIPeR and ETHZ datasets, since images of VIPeR and ETHZ have higher resolution than those of CAVIAR4REID. eSDC_knn adopts high dimension feature vectors of dense SIFT and dense color histogram to describe each image in detail. Moreover, except the 3% descent at rank 1 matching rate, our proposed algorithm achieves an approximate matching rate with eSDC_knn on the average.

It is worth noting that the performance of our algorithm can be elevated by adding other complementary features. The rank 1 matching rate can be improved by average 6% when our algorithm is combined with SDC and MSCR.

Moreover, the performance of our algorithm is further compared with two similar algorithms, namely Wu et al. [3] and Yang et al. [4], which have the similar idea with our work in the salient feature representation. Note that our algorithm addresses the pose variation and low resolutions. While Wu et al. [3] adopt pose priors(labeled as '*PP*') to improve the robustness against viewpoint or pose changes, learn person-specific discriminative features(labeled as '*DF*') to boost the performance, and use the SVM as the metric learning method. Yang et al. [4] propose a salient color name based color descriptor (*SCNCD*), based on *SCNCD*, color distributions in four different color spaces (RGB, normalized rgb, l1l2l3 and HSV) are calculated and fused to address the illumination problem, and finally, a metric learning method is utilized to distance metrics.

The experimental results are shown in Table 5. It is obvious that our algorithm outperform Wu et al. [3] in all three cases, namely the pose prior algorithm (*SVM* + *PP* [3]), the subject-discriminative feature selection(*SVM* + *DF* [3]), and their combination (*SVM* + *PP* + *DF* [3]). Specially, our algorithm yields 8% improved matching rate on rank 1 when compared with *SVM* + *PP* + *DF* [3].

When compared with Yang et al. [4], our algorithm can achieve almost 7.7% improvement on rank 1 matching rate when a single color space is adopted ($SCNCD_{RGB}$ [4], the *SCNCD* based on RGB color space). Note that $SCNCD_{RGB}$ [4] can achieve the best performance

Table 4
Comparison of top r rank matching rate (%).

Methods		CAVIAR4REID	VIPeR	ETHZ SEQ1	ETHZ SEQ2	ETHZ SEQ3	Avg.
SDALF [1]	$r = 1$	16.51	18.35	63.45	64.5	71.38	*46.83*
	$r = 5$	31.39	37.34	83.28	85.34	91.50	*65.77*
	$r = 10$	41.64	50.63	89.30	88.60	95.45	*73.12*
	$r = 20$	56.24	63.92	94.89	94.59	98.80	*81.68*
eSDC_knn [6]	$r = 1$	20.45	25.63	82.41	76.61	89.86	*58.99*
	$r = 5$	35.95	45.89	90.96	89.66	96.77	*71.84*
	$r = 10$	45.81	59.49	94.22	94.47	98.26	*78.45*
	$r = 20$	58.92	73.42	97.33	98.50	99.82	*85.59*
RbFS (our)	$r = 1$	26.56	22.47	80.41	73.88	75.93	*55.85*
	$r = 5$	41.34	46.84	91.83	87.40	91.38	*71.75*
	$r = 10$	50.42	56.65	94.65	93.90	96.59	*78.44*
	$r = 20$	62.28	73.73	97.17	96.24	98.39	*85.56*
RbFS and MSCR and SDC	$r = 1$	32.70	28.48	92.24	79.29	89.39	***64.42***
	$r = 5$	48.90	50.63	97.28	91.33	97.00	***77.02***
	$r = 10$	57.90	64.56	98.43	95.34	98.91	***83.02***
	$r = 20$	69.51	76.27	99.25	98.49	99.98	***88.7***

Table 5
Comparison of top *r* rank matching rate (%) on the VIPeR dataset.

Rank	$SCNCD_{RGB}$ [4]	$SCNCD_{all}$ [4]	*SVM + DF* [3]	*SVM + PP* [3]	*SVM + PP + DF* [3]	Our
r = 1	20.7	33.7	18.7	19.4	21.4	28.48
r = 5	47.2	62.7	42.2	44.0	45.9	50.63
r = 10	60.6	74.8	56.5	59.0	60.5	64.56
r = 20	75.1	81.3	73.4	74.3	75.9	76.27

when compared with *SCNCD* based on other three color spaces (i.e., HSV, normalized rgb, and l1l2l3). However, $SCNCD_{all}$ [4] can achieve 33.7% recognition accuracy on rank 1, which means that *SCNCD* presents certain illumination invariance when features are fused under the above four color space. Since the color feature is extracted only in the HSV color space in our algorithm, $SCNCD_{all}$ [4] outperforms our algorithm under this situation. This result implies that single color space is hardly to solve all kinds of illumination problems, and multiple compensative color models should be fused to improve the robustness against the illumination changes. Therefore, we will consider the features representation in multiple complementary color spaces for coping with the illumination changes in our further work.

It should be noted that the metric learning is adopted to improve person re-identification performance in the literatures [3,4]. This point is further proved by the research of Zhao et al., SalMatching [7] proposes an supervised salience matching based on *eSDC* [6], which improve the rank 1 matching rate by 4% compared with *eSDC* [6]. Metric learning is unemployed in our algorithm. We will explore the robust feature representation while metric learning is required in our future work.

6. Conclusion

In this paper, we propose a person re-identification algorithm by exploiting region-based feature salience. There are two main contributions: First, a part-based feature extraction is adopted to represent different parts with different features according to the characteristics of each part, and then each kind of features are separately represented in the multi-feature fusion to make full use of its salience and retain its intrinsic meanings. Secondly, a new salient color descriptor (SCD) is extracted to embody the color representation and discrimination through the salient patch detection and representative color feature extraction for the salient patch. The experiment shows that our algorithm yields improved matching rate for the images of low resolution and obvious pose variation, the approximate matching rate as eSDC_knn is achieved for the images of moderate resolution and serious view changes. Since our algorithm mainly embodies the availability and salience of extracted feature for each region. The rank 1 matching rate can be improved by average 6% when MSCR and SDC is complementary to our algorithm, and robustness to pose, viewpoint and illumination variations is achieved. The proposed algorithm can improve the recognition performance comparing with the state-of-the-art algorithms.

The proposed algorithm mainly focuses on the exploration of the feature salience. Although it is showed to be robust to pose variation and low resolutions, features mentioned in our algorithm are not enough to describe each part in the complex practical surveillance system. More robust features should be extracted to represent the appearance. Moreover, for the metric distance fusion of upper part and lower part, the importance of the same part is influenced by occlusion and severe pose change across different views. Adaptively adjusting the weights of different parts should be employed to further improve the accuracy of person re-identification. The above two points will be our ongoing works.

References

[1] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 2360–2367.

[2] O. Oreifej, R. Mehran, M. Shah, Human identity recognition in aerial images, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 709–716.

[3] Z. Wu, Y. Li, R. Radke, Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) (2014) 1.

[4] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, Stan Z. Li, Salient color names for person re-identification, Eur. Conf. Comput. Vision (ECCV) 8689 (2014) 536–551.

[5] L. Ma, X. Yang, Y. Xu, J. Zhu, Human identification using body prior and generalized EMD, IEEE Conf. Image Process. (ICIP) (2011) 1441–1444.

[6] 4.R. Zhao, W. Ouyang, X. Wang, Unsupervised salience learning for person re-identification, IEEE Conf. Comput. Vision Pattern Recognit. (CVPR) (2013) 3586–3593.

[7] R. Zhao, W Ouyang X. Wang, Person re-identification by salience matching, in: IEEE International Conference on Computer Vision (ICCV), 2013.

[8] L. Bazzani, M. Cristani, V. Murino, SDALF: modeling human appearance with symmetry-driven accumulation of local features, in: Person Re-Identification, Springer, 2014, pp. 43–69.

[9] L. Bazzani, M. Cristani, A. Perina, V. Murino, Multiple-shot person re-identification by chromatic and epitomic analyses, Pattern Recognit. Lett. (PRL) (2012) 898–903.

[10] L. Bazzani, M. Cristani, V. Murino, Symmetry-driven accumulation of local features for human characterization and re-identification, Comput. Vis. Image Underst. (2013) 130–144.

[11] W.R. Schwartz, L.S. Davis, Learning discriminative appearance-based models using partial least squares, in: 2009 XXII Brazilian Symposium on Proc. Computer Graphics and Image Processing (SIBGRAPI), IEEE, 2009, pp. 322–329.

[12] Y. Cai, M. Pietikäinen, Person re-identification based on global color context, in: Proc. Computer Vision–ACCV 2010 Workshops, Springer, 2011, pp. 205–215.

[13] A. Bedagkar-Gala, S.K. Shah, Multiple person re-identification using part based spatio-temporal color appearance model, in: IEEE International Conference on Computer Vision (ICCV), 2011, pp. 1721–1728.

[14] M. Piccardi, E.D. Cheng, Multi-frame moving object track matching based on an incremental major color spectrum histogram matching algorithm, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 19.

[15] B. Prosser, S. Gong, T. Xiang, Multi-camera matching under illumination change over time, in: Proc. Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SFA2 2008, 2008.

[16] B. Prosser, S. Gong, T. Xiang, Multi-camera matching using bi-directional cumulative brightness transfer functions, in: Proc. BMVC, Citeseer, 2008, pp. 161–164.

[17] K. Jeong, C. Jaynes, Object matching in disjoint cameras using a color transfer approach, Mach. Vis. Appl. (2008) 443–455.

[18] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discriminant analysis for pedestrian re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3318–3325.

[19] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: Computer Vision–ECCV 2008, Springer, 2008, pp. 262–275.

[20] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, Information-theoretic metric learning, ACM Int. Conf. Mach. Learn. (2007) 209–216.

[21] F. Xiong, M. Gou, O. Camps, M. Sznaier, Person re-identification using kernel-based metric learning methods, in: European Conference on Computer Vision (ECCV), 2014.

[22] X. Liu, M. Song, D. Tao, X. Zhou, C. Chen, J. Bu, Semi-supervised coupled dictionary learning for person re-identification, in: IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2014, pp. 3550–3557.

[23] A. Globerson, S. Roweis, Metric learning by collapsing classes, Proc. Nips (2005) 451–458.

[24] J. Lee, R. Jin, A.K. Jain, Rank-based distance metric learning: an application to image retrieval, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.

[25] M. Schultz, T. Joachims, Learning a distance metric from relative comparisons, Proc. NIPS (2003).

[26] L. Yang, R. Jin, R. Sukthankar, Y. Liu, An efficient algorithm for local distance metric learning, Proc. AAAI (2006).

[27] E.P. Xing, A.Y. Ng, M.I. Jordan, S. Russell, Distance metric learning with application to clustering with side-information, Adv. Neural Inf. Process. Syst. (2003) 521–528.

[28] K. Weinberger, J. Blitzer, L. Saul, Distance metric learning for large margin nearest neighbor classification, Adv. Neural Inf. Process. Syst. (2006) 1473.

[29] W. Zheng, S. Gong, T. Xiang, Person re-identification by probabilistic relative distance comparison, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 649–656.

[31] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 886–893.

[32] A. Ess, B. Leibe, K. Schindler, L. Van Gool, A mobile vision system for robust multi-person tracking, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.

[33] A. Ess, B. Leibe, L. Van Gool, Depth and appearance for mobile scene analysis, in: IEEE International Conference on Computer Vision (ICCV), 2007, pp. 1–8.

[34] C. Liu, S. Gong, C.C. Loy, X. Lin, Person re-identification: what features are important?, ECCV Re-id Workshop (2012) 391–401

[36] M. Hu, W. Hu, T. Tan, Tracking people through occlusions, in: IEEE Conference on Pattern Recognition (ICPR), 2004, pp. 724–727.

[37] Y. Rubner, C. Tomasi, L.J. Guibas, The earth mover's distance as a metric for image retrieval, Int. J. Comput. Vision (2000) 99–121.

[38] J. Van De Weijer, C. Schmid, Coloring local feature extraction, in: Computer Vision–ECCV 2006, Springer, 2006, pp. 334–348.

[39] W. Zheng, S. Gong, T. Xiang, Associating groups of people, British Machine Vision Conference (BMVC), 2009, pp. 23.1–23.11.

[40] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 24 (2002) 971–987.

[41] Y. Geng, H. Hu, J. Zheng, B. Li, A person re-identification algorithm by using region-based feature selection and feature fusion, in: IEEE Conference on Image Processing (ICIP), 2013, pp. 3363–3366.

[42] G. Zeng, H. Hu, Y. Geng, A person re-identification algorithm based on color topology, in: IEEE Conference on Image Processing (ICIP), 2014, pp. 2447–2451.