Pattern Recognition 71 (2017) 218–234

Contents lists available at ScienceDirect

Pattern Recognition

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

Multiple metric learning based on bar-shape descriptor for person re-identification

Cairong Zhao[a,b,*], Xuekuan Wang[a,*], Wai Keung Wong[b,e], Weishi Zheng[d], Jian Yang[c], Duoqian Miao[a,*]

[a] Department of Computer Science and Technology, Tongji University, Shanghai, China
[b] Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong
[c] Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information (NJUST), Ministry of Education, Nanjing, China
[d] School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China
[e] Hong Kong Polytechnic University Shenzhen Research Institute, China

ARTICLE INFO

Article history:
Received 4 June 2016
Revised 15 May 2017
Accepted 7 June 2017
Available online 8 June 2017

Keywords:
Person re-identification
Multiple bar-shape descriptor
Multiple metric learning

ABSTRACT

The robust structural feature extraction and similarity measure play critical roles in person re-identification. This paper presents a novel algorithm named Multiple Metric Learning based on Bar-shape Descriptor (**MMLBD**) for person re-identification. Specifically, we first propose a new Multiple Bar-shape Descriptor that can take full account of the spatial correlation between the center points and their adjacent points on different directions. It captures further histogram features based on a novel color difference weight factors with an overlapping sliding window, which can depict the local variations and consistency in the whole image. The similarity and dissimilarity of samples are used to train the weight factor of features and an optimal subspace could be obtained at the same time. Next, we provide an effective multiple metric learning method fusing two-channel bar-shape structural features via the optimal similarity pairwise measure obtained by a dissimilarity matrix. This measure can fully mine the discriminative information and eliminate redundancy in the similar features, which make the **MMLBD** simple and effective. Finally, evaluation experiments on the i_LIDS, CAVIAR4REID and WARD data-sets are carried out, which compare the proposed **MMLBD** with the corresponding methods. Experimental results demonstrate that the **MMLBD** is more effective and robust against visual appearance variations.

1. Introduction

Person re-identification (**Re-ID**) is described for matching same pedestrians across disjoint camera views in a multi-camera system, and is increasingly receiving attention as a key component of video surveillance [1]. The task of person Re-ID is to recognize the occurrence of a target pedestrian captured by one camera from a gallery of labeled subject. Recently, various descriptors based on pedestrian's appearance have been developed. However, it is still difficult to extract robust and discriminative features from the appearance of pedestrians, due to complexity of the environment that is affected by the changes of illumination, pose, viewpoint, occlusion, image resolution and camera setting in the non-overlapping camera systems [2]. At present, the state-of-the-art approaches for person Re-ID are mainly divided into two groups: (1) the appearance-based approach which designs distinctive and effective descriptors to represent a person's appearance; (2) the metric learning approach which learns a suitable measure to minimize the similarity between the same people and maximize the similarity between the different people. The main developments of person Re-ID are shown in Table 1.

Over the past few years, low-level features such as color [3–5] and texture [6,7], have been widely applied to appearance-based representation. Furthermore, some studies including bag-of-words model [8], local maximal occurrence (LOMO) [9], hierarchical Gaussian descriptor [10], recurrent feature aggregation network (RFA-Net) [11], and hash feature [12, 43], etc., have attempted to integrate them to capture more robust and reliable features. Apart from these methods, the deep learning [13–15] is especially noteworthy model which has exhibited an excellent performance in learning representation for person Re-ID. Unfortunately, it is still extremely difficult to extract a stable feature representation which can effectively adapt to severe changes and misalignment across disjoint views. Besides, neither color nor texture features are able

* Corresponding author. Department of Computer Science and Technology, Tongji University, Shanghai 201804, China.
E-mail addresses: zhaocairong@tongji.edu.cn (C. Zhao), wxktongji@163.com (X. Wang), dqmiao@tongji.edu.cn (D. Miao).

http://dx.doi.org/10.1016/j.patcog.2017.06.011
0031-3203/

Table 1
Main developments of person Re-ID.

Authors	Year	Approaches	Structural information	Remark
Gray and Tao[47].	2008	ELF	No	Appearance
Farenzena et al. [6]	2010	SDALF	Yes	Appearance Metric learning
Avraham et al. [49]	2012	Transfer	Yes	Transfer learning
Zheng et al. [22]	2013	RDC	No	Appearance + Metric learning
Zhao et al. [18]	2013	Salience	Yes	Appearance + Matching
Pedagadi et al. [23]	2013	LFDA	No	Metric learning
Xiong et al. [24]	2014	Kernel	Yes	Metric learning
Yang et al. [3]	2014	Color Name	No	Appearance
Ma et al [26]	2014	Multiple tasks	No	Metric learning
Shen et al. [20]	2015	Structure	Yes	Structure learning
Lisanti et al. [27]	2015	Sparse Rank	No	Rank learning
Ahmed et al. [29]	2015	Deep	Yes	Appearance + Metric learning
Liao et al. [9]	2015	LOMO	Yes	Appearance + Metric learning
Matsukawa et al. [10]	2016	GOG	Yes	Appearance
Tao et al. [32]	2016	DR-KISS	No	Metric learning
Zheng et al. [33]	2016	Transfer	Yes	Transfer learning

to describe the structural shape characteristics of pedestrians exactly.

Different from color and texture features, the structural features captured the local shape information from images, as they focus on the spatial correlation between points retaining the color and texture information [16]. M. Farenzena, et al. [6] designed symmetry-driven accumulation of local features (**SDALF**) to capture multiple varieties of information from three stable parts of human body based on the maximally stable color regions (**MSER**) [17]. By contrast, R. Zhao, et al. [18] learned human salience in an unsupervised manner to find reliable and discriminative matched patches for person Re-ID and S. Iodice, et al. [19] utilized symmetry principles, as well as structural relations among salient features to obtain structure information via a graph matching method. Besides, Y. Shen, et al. [20] integrated a global matching constraint over the learned correspondence structure to exclude cross-view misalignments during the image patch matching process. Metric learning is another interesting aspect of the person Re-ID. Generally, the existing metric models could be divided roughly into two categories: non-learning and learning methods. Many of the models simply choose a standard distance such as $l_{1,2}$–norm [21]. However, they treat all features equally instead of discarding bad features selectively. Thus the matching results are always undesirable. On the contrary, the metric learning based measurement approaches, including Relative Distance Comparison (RDC) [22], Local Fisher Discriminant Analysis (LFDA) [23] Kernel-Based Metric [24], Mahalanobis Distance Learning [25], Multi-task Distance Metric Learning [26], Iterative Re-Weighted Sparse Ranking [27], Multiple Metric Learning [28], Deep Metric Learning [29], Cross-view Quadratic Discriminant Analysis (XQDA) [9], Tensor Learning [30], Saliency Learning Model [31], Dual-Regularized KISS (DR-KISS) [32] and Transfer Learning Model [33, 49], etc., learn typically a discriminative similarity between the same and different persons across camera pairs. Although these metric learning methods outperform the existing person Re-ID benchmarks, they are still limited by some classical problems, such as robust feature representation and small sample size (***SSS***) for model learning.

To address this problem, we put forward a Multiple Bar-shape Descriptor (**MBD**) which takes advantage of a hybrid encoding strategy combining the color granularity and local binary encoding form bar-shape structures, shown in Fig. 2, to capture the robust structural information. Differently, we apply ***Color Difference Weight, Overlapping Slide Window*** and ***Max-pooling Operator*** to consider more visual information and ensure more local structural information. Meanwhile, local encoding histograms are captured from two channels with multiple orientations to ensure the low dimensionality of feature descriptor and the robustness of the changes of illumination. Then, the discriminant weight subspace learning are utilized for the Canberra distance. Furthermore, we propose a novel relative distance fusing algorithm to integrate multi-orientation bar-shape structural features. Instead of learning a metric over hand-crafted features, we utilize the similarity of metric to extract optimal pairwise distance, based on dissimilarity matrix, and fuse multiple distances for person Re-ID. It can avoid complex model learning effectively. The main contributions are highlighted as follows:

(1) We design a novel ***Multiple Bar-shape Descriptor (MBD)*** which applies a hybrid encoding strategy to extract bar-shape structural features, integrating multi-channel local binary pattern and color granularity encoding.
(2) We present a new metric learning method based on the similarity of distances and fuse multiple metrics via optimal relative distance pairs to learn a robust distance function dealing with the complex matching fusion problem. Meanwhile, we put forward an effective color difference weight factor based on the similarity and dissimilarity of samples to characterize different important attributes of different features.
(3) Experimental results show the proposed method of **MMLBD** is more effective and robust against visual appearance variations, achieving superior performance on three public person Re-ID data-sets in most cases.

The remainder of this paper is organized as follows. We review the related works and introduce the theory of the proposed approach in Section 2 and Section 3, respectively. Then, we carry out the comparative experiments on three public person Re-ID data-sets and give the detailed discussions based on the experimental results in Section 4. Finally, conclusions are made in Section 5.

2. Related work

This paper aims to seek an effective method for person Re-ID based on multi-channel feature extraction. Firstly, we present an overview of the relevant works, i.e., Census Transform Pyramid [34,35] and binary interaction mechanism [35].

2.1. Census transform pyramid

A representative of structural image descriptors is Local Binary Pattern (LBP) proposed firstly by Ojala et al. [36] as a gray-scale invariant texture descriptor. The LBP code is obtained by its circularly symmetric n-neighbors in a circle of radius r with the pixel value of the central point and arranging the results as a binary string. It is robust for the changes of illumination. Based on this,

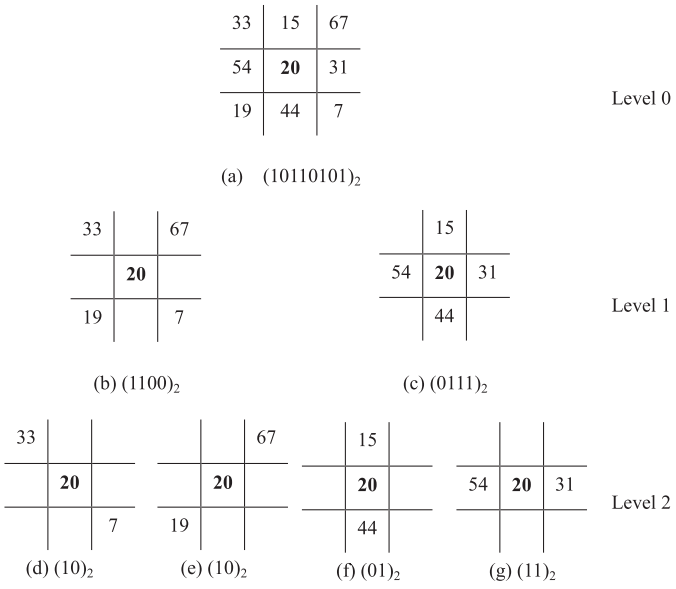


Fig. 1. Census transform pyramid.

many variants are proposed, such as CLBP [37], CBP [38], LDP [39], LTP [40], LQP [41], CS-LBP [42], et al. However, these approaches cannot address the curse of dimensionality with the increase of points being encoded. In mCENTRIST [35], Y. Xiao, et al. utilize the Census Transform (CT) pyramid to reduce the dimensionality, achieving higher accuracy rate for scene categorization. The Census Transform pyramid is shown in Fig. 1 and comprised of three levels in which level 0 is standard CT that represents a center point with binary coding by its 8-adjacency points from top-right to bottom-left. Compared with the center point, a value greater than the pixel value is defined as 1, otherwise 0. Then we can obtain the binary coding:

$$CT_1 = (10110101)_2 \tag{1}$$

However, the CT_1 only represents the binary code of one channel. For color images encoded by level 0, it should be described as follows:

$$\overrightarrow{mCT_{level0}} = \underbrace{(CT_1 CT_2, \ldots, CT_n)}_{n-channels} \tag{2}$$

CT_1, CT_2 and CT_3 describe the binary coding with three channels of color images respectively. It could also include four channels which consist of three color channels and one ***Sobel*** operator. For level 1 and 2, the level 0 is divided into two or four sub-mCTs on center directions. The approach of binary coding is the same as level 0. For level 1, two binary coding with two directions can be obtained:

$$CT_1^1 = (1100)_2 \tag{3}$$

and

$$CT_1^2 = (0111)_2 \tag{4}$$

Similarly, to capture multi-channel information of color images, the multiple sub-mCT value with different directions is defined as follows:

$$\overrightarrow{mCT^1_{level1}} = \underbrace{\left(CT_1^1, CT_2^1, \ldots, CT_n^1\right)}_{n-channels} \tag{5}$$

and

$$\overrightarrow{mCT^2_{level1}} = \underbrace{\left(CT_1^1, CT_2^1, \ldots, CT_n^1\right)}_{n-channels} \tag{6}$$

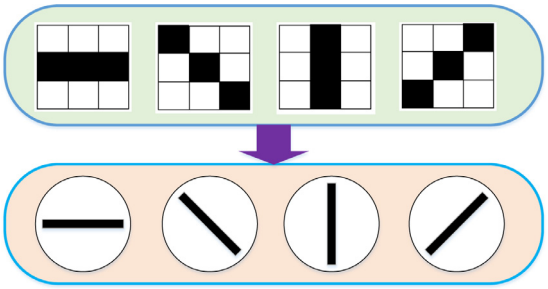


Fig. 2. Bar-shape structures with four direction-senses about 0°, 45°, 90° and 135°.

In this way, compared with level 0, the dimensionality of feature vector is reduced from $2^{8\times n}$ to $m \times 2^{k\times n}$ where n is the number of channels, m is the number of directions and k is the length of binary coding on different directions.

2.2. Binary interaction mechanism

For real application, the feature descriptor extracted via sub-mCT is also impractical if n is higher. To address this problem, the binary interaction mechanism is proposed [35] and it would consider a group of two-channel sub-mCT histograms rather than all channels directly to avoid prohibitively huge dimensionality. Let $C = \{c_1, c_2, c_3, c_4\}$ be a four-channel color image and six channel pairs could be derived based on binary interaction mechanism, defined as $c'_1 = \{c_1, c_2\}$, $c'_2 = \{c_1, c_3\}$, $c'_3 = \{c_1, c_4\}$, $c'_4 = \{c_2, c_3\}$, $c'_5 = \{c_2, c_4\}$, $c'_6 = \{c_3, c_4\}$. Then, two-channel sub-mCT histogram pairs will be extracted from each c'_i respectively and defined as $C' = \{c'_1, c'_2, c'_3, c'_4, c'_5, c'_6\}$. Through the binary interaction mechanism, the dimensionality of feature descriptor is reduced form $m \times 2^{k\times n}$ to $\binom{n}{2} \times m \times 2^{k\times 2}$ where n is the number of channels, m is the number of directions and k is the length of local binary coding on different directions. And it achieves a balance between computational efficiency and discriminative power.

3. The proposed algorithm (MMLBD)

It is a well-known fact that person Re-ID is a challenging problem because of big intra-class variations in illumination, pose, viewpoint, and occlusion. However, the appearance of a person is usually rich in stable bar-shape structures, as shown in Fig. 2, which have significant direction-senses about 0°, 45°, 90° and 135°, respectively in human vision system. These bar-shape structures are concerned about the spatial correlation of adjacent points. Moreover, they are robust for the changes of illumination, rotation and translation. In addition, ***Sobel*** operator can be used to find the approximate absolute gradient magnitude at each point in an input image [35].

In this paper, we take advantage of a novel hybrid encoding strategy which consists of color difference weight histograms based on Census Transform Pyramid and color granularity encoding with four color channels ($L^*a^*b^*$ and ***Sobel***). Color, texture, color difference and spatial structure information are all considered in our approach. To ensure lower dimensionality, we apply binary interaction mechanism to capture multiple two-channel descriptors with four orientations. Furthermore, we project descriptors into an optimal subspace and take advantage of an overlapping sliding window to extract feature histogram from local contrast-normalized cells, eliminating the local variations and enhancing the adaptability to illumination variations, shadowing and small shift in images. Meanwhile, we make full use of the similarity of dissimilar samples to evaluate the qualities of different

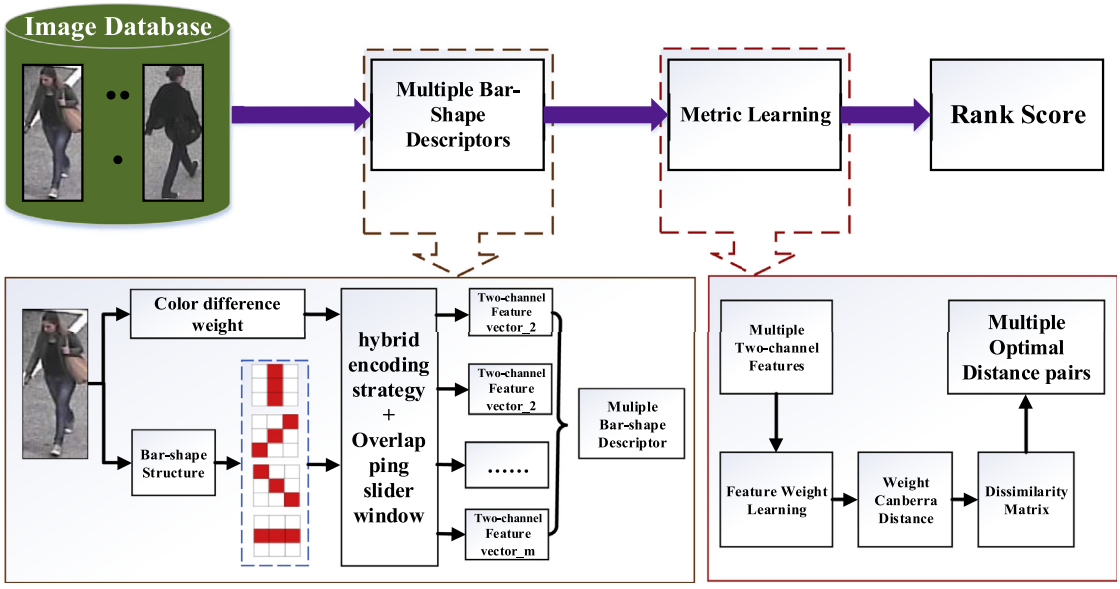


Fig. 3. The process of the proposed approach.

features by statistical learning and fuse more discrimination information to improve the matching rates of person Re-ID.

After the process of feature extraction, we can obtain multiple bar-shape features. Considering the similarity of same person, we make full use of the comparison of relative distances to learn multiple optimal distance pairs based on dissimilarity matrix and fuse them with their credibility defined as the weight factor to achieve a more effective performance. The process of our proposed approach is shown in Fig. 3 and be introduced in detail as follows.

3.1. Multiple Bar-Shape Descriptor

In order to capture more useful information, we focus on multiple channels of a color image, rather than a gray image. Based on some existing studies, the color space of $L^*a^*b^*$ is more suitable for perception of color in human vision system [44]. Hence, the color space of $L^*a^*b^*$ is adapted to person Re-ID. Meanwhile, we design a **multiple bar-shape descriptor (MBD)** which takes advantage of a hybrid encoding strategy with the consideration of the color difference weight, Census Transform pyramid, color granularity encoding and binary interaction mechanism. It can capture effectively the bar-shape structures on person's appearance. Then, we take advantage of an overlapping sliding window to extract the histogram based on hybrid encoding and color difference weight. Considering the salient features, we capture local histogram by Max-pooling operator on the horizontal direction. The process of multiple bar-shape descriptor is shown in Fig. 4.

(A). Hybrid encoding strategy

In our approach, the hybrid encoding strategy consists of sub-mCT and color granularity encoding with multiple channels and multiple orientations, as shown in Fig. 5. For sub-mCT, we denote the local binary encoding of pixel A with different channels and orientations as $BSL^c_\theta(A)$. For color granularity encoding, we granular the color space of $L^*a^*b^*$ into 4, 4, 4 bins and the Sobel channel into 9 bins, and denote it as $BSC^c_\theta(A)$. Then, the $BSL^c_\theta(A)$ and $BSC^c_\theta(A)$ are defined as follows:

$$BSL^c_\theta(A) = \sum_{i=0}^{N-1} s\left(g^c_{\theta,i} - g^c_{center}\right) \times 2^i \tag{7}$$

$$BSC^c_\theta(A) = \sum_{i=0}^{N-1} (k_c)^i \times g^c_{\theta,i} \tag{8}$$

$$s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \tag{9}$$

where g^c_{center} is the pixel value of point A in channel-c, N is the number of the neighbor points of center point A on direction-θ, $g^c_{\theta,i}$ is the pixel value of point A's i-th neighbor on the direction-θ, k_c is the number of bins in channel-c, c is one of the L^*, a^*, b^*, *sobel* and θ is the direction (0°, 45°, 90° and 135°). Considering level two for sub-CTs on center directions (More details can be found in Section 2.1), the value of N is 2.

Furthermore, we utilize binary interaction mechanism [35] and choose two-channel pairs from L^*, a^*, b^* and *Sobel* to obtain two-channel bar-shape structures on different directions, defined as

$$BSL^{i,j}_\theta(A) = BSL^i_\theta(A) \times 2^N + BSL^j_\theta(A) \tag{10}$$

$$BSC^{i,j}_\theta(A) = BSC^i_\theta(A) \times k_j + BSC^j_\theta(A) \tag{11}$$

where i, j represent two different channels, $BSL^{i,j}_\theta(A)$ is two-channel local binary encoding on direction-θ, N is the number of neighbors on direction-θ, $BSC^{i,j}_\theta(A)$ is two-channel granularity encoding with direction-θ, k_j is the number of bins with channel-j. In our approach, for a four-channel image, 6 two-channel pairs can be derived by binary interaction on different directions and defined as

$$BL_\theta = \left\{BSL^{L,a}_\theta, BSL^{L,a}_\theta, BSL^{L,sobel}_\theta, BSL^{a,b}_\theta, BSL^{a,sobel}_\theta, BSL^{b,sobel}_\theta\right\} \tag{12}$$

and

$$BC_\theta = \left\{BSC^{L,a}_\theta, BSC^{L,b}_\theta, BSC^{L,sobel}_\theta, BSC^{a,b}_\theta, BSC^{a,sobel}_\theta, BSC^{b,sobel}_\theta\right\} \tag{13}$$

For different directions, we can obtain two sets of multiple bar-shape feature images, denoted as $BL = \{BL_\theta\}$, $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ and $BC = \{BC_\theta\}$, $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$.

By using binary interactions, the dimensionality of feature is only $\binom{n}{2} \times (Dim(BSL^c_\theta) + Dim(BSC^c_\theta)) \times m$ described in Section 2.1 for n-channel images and it avoids prohibitively huge features,

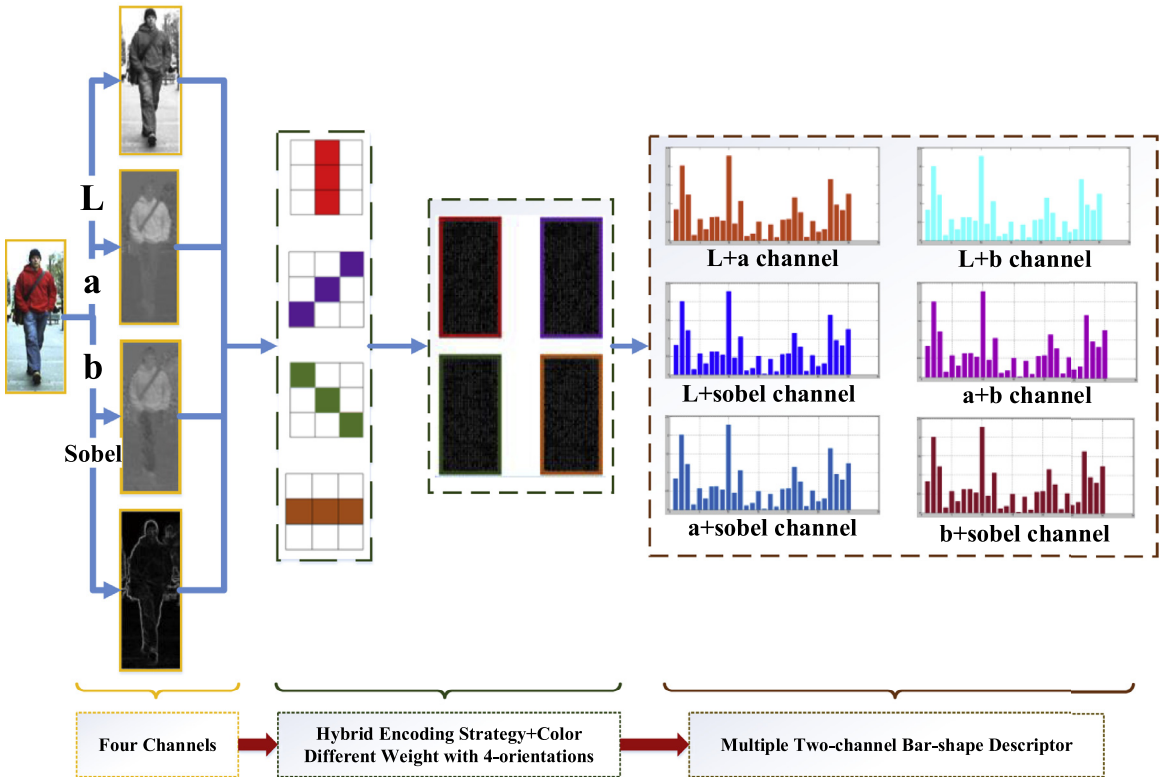


Fig. 4. The process of multiple bar-shape descriptor.

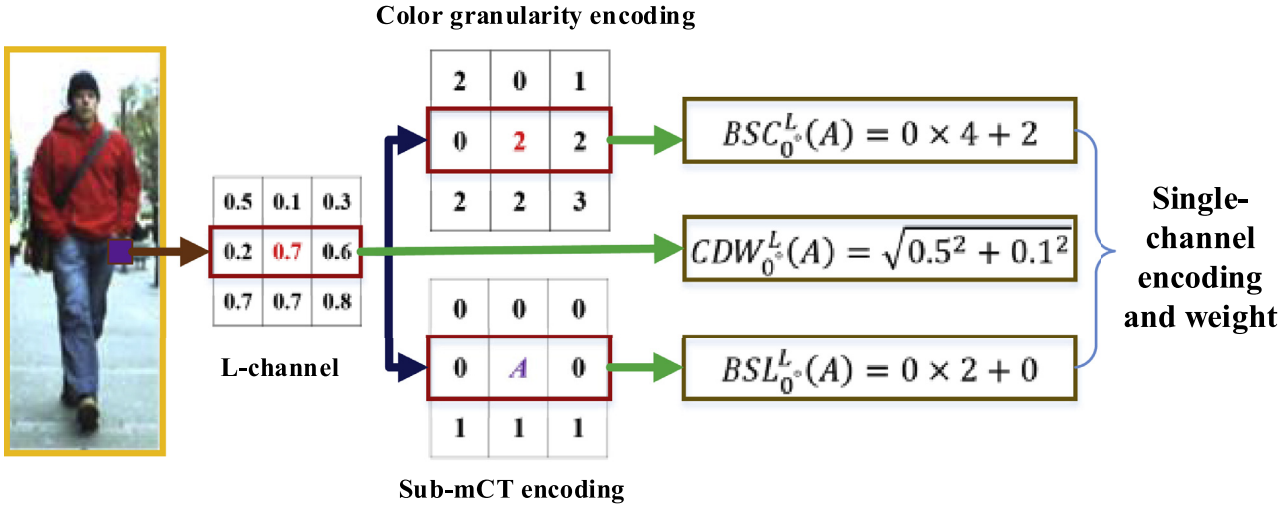


Fig. 5. The processing of sub-mCT and color granularity encoding with single-channel and single-orientation ($c = L$ and $\theta = 0^\circ$).

where n is the number of channels, m is the number of directions and we let n, m to 4 empirically in our approach.

(B). Color difference weight

Meanwhile, in order to fuse the color difference information, we obtain the color difference weight of $CDW_\theta^c(A)$ for each original pixel $A(x,y)$ with single channel. It is more invariant to monotonic changes of illumination and defined as follows:

$$CDW_\theta^c(A) = \sum_{i=0}^{N-1} \left| g_{\theta,i}^c - g_{center}^c \right| \tag{14}$$

Considering that the center point is more important than the surrounding points in a person image [44], we define the weight map ($w(A)$) for each point and the color difference weight with two-channel bar-shape structure as $CDW_\theta^{c_1,c_2}(A)$:

$$CDW_\theta^{c_1,c_2}(A) = w(A) \times \sqrt{CDW_\theta^{c_1}(A)^2 + CDW_\theta^{c_2}(A)^2} \tag{15}$$

$$w(A) = e^{-\frac{(x-\mu_x)^2}{2\sigma_x}} \times e^{-\frac{(y-\mu_y)^2}{2\sigma_y}} \tag{16}$$

where c_1, c_2 denote the different channels, $\mu_x = L_x/2$, $\mu_y = L_y/2$, $\sigma_x = L_x/4$ and $\sigma_y = L_y/4$. In Eq. (16), x denotes the row number of the image matrix and y denotes the column number of the image matrix. L_x and L_y denote the image height and width, respectively.

For a four-channel person image, we can obtain six two-channel color difference weight maps, defined as $W_\theta = \{CDW_\theta^{L,a}, CDW_\theta^{L,b}, CDW_\theta^{L,sobel}, CDW_\theta^{a,b}, CDW_\theta^{a,sobel}, CDW_\theta^{b,sobel}\}$.

(C). Histogram extraction with overlapping slide window

In order to further enhance the robustness to illumination, we capture statistical histogram features to describe the characteristics of multiple bar-shape feature images from two-channel hybrid encoding maps including $BL_\theta(A)$ and $BC_\theta(A)$ with an overlapping sliding window with size of $n \times n$, and the weight is obtained by $W_\theta(A)$. For a point $A = (x, y)$, its two-channel local binary encoding is $BSL_\theta^{i,j}(A)$ and color granularity encoding is $BSC_\theta^{i,j}(A)$. Besides, the color difference weight of point A is defined as $CDW_\theta^{i,j}(A)$. Then, the $BSL_\theta^{i,j}(A)$-th value of the histogram $HL_\theta^{i,j}$ is computed as follows:

$$HL_\theta^{i,j}\left(BSL_\theta^{i,j}(A)\right) = \sum_{A\in R} CDW_\theta^{i,j}(A) \tag{17}$$

and the $BSC_\theta^{i,j}(A)$-th value of the histogram $HC_\theta^{i,j}$ is computed as follows:

$$HC_\theta^{i,j}\left(BSC_\theta^{i,j}(A)\right) = \sum_{A\in R} CDW_\theta^{i,j}(A) \tag{18}$$

where R is the region of the sliding window in an image, θ is the orientation and i, j represent two different channels. In order to better characterize invariance to illumination, shadowing, etc., we contrast-normalize the local histograms captured from regions of the overlapping sliding window. Then, we can obtain the multiple two-channel bar-shape structural histograms, defined as $H_{i,j} = [HL_{0^\circ}^{i,j}, HC_{0^\circ}^{i,j}, HL_{45^\circ}^{i,j}, HC_{45^\circ}^{i,j}, HL_{90^\circ}^{i,j}, HC_{90^\circ}^{i,j}, HL_{135^\circ}^{i,j}, HC_{135^\circ}^{i,j}]$, $i, j \in \{L, a, b, sobel\}$. Moreover, we take advantage of max-pooling to capture the multi-scale bar-shape structural feature from multi-scale person image [9].

The bar-shape structural color difference weight histogram fuses the color, color difference, texture information and the spatial correlation of points with multiple channels and orientations. It can represent effectively the structural appearance of a person and make full use of contrast-normalization to enhance the adaptability to illumination variations, shadow, etc. Hence, these descriptors are relatively suitable to represent the appearance of human in person Re-ID.

3.2. Multiple optimal distance pairs metric learning

(A). Discriminant weight subspace distance

Owing to the strategy of overlapping, the descriptor of ***MBD*** is high-dimensional, redundant and sparse. Therefore, we apply subspace projection to seek a low-dimensional feature subspace instead of the original feature. This problem is solved by local fisher discriminant analysis (***LFDA***) which is required to solve a generalized eigenvalue problem of very large scatter $d \times d$ matrices [23]. Meanwhile, in order to reduce the bad features negative effects on the metric, we utilize a discriminant weight factor to evaluate the quality of feature. Intuitively, it hopes the distance to be closer for intra-class feature pairs, and further for inter-class pairs. Thus, we define intra-class and inter-class distances (Eq. 19–21) to estimate the similarity of same person and dissimilarity of different person respectively. Then, we capture inherent discriminant information to describe the importance of feature. In detail, the discriminant weight of feature is computed by

$$W_f = e^{\sum_D - \sum_S} \tag{19}$$

$$\sum_D = \sum_{x_i, x_j \in D} log\left|x_i - x_j\right| \tag{20}$$

$$\sum_S = \sum_{x_i, x_j \in S} log\left|x_i - x_j\right| \tag{21}$$

where Σ_S, Σ_D denote the distance vectors for similar pairs and dissimilar pairs respectively and x_i, x_j represent two different samples. Then, for different two-channel bar-shape descriptors, the discriminant weight of feature $W_{f_{c_1,c_2}}$ can be learned easily from the training samples, where c_1, $c_2 \in \{L, a, b, sobel\}$ represent different channels. Then, we apply Canberra distance [44], combining with the weights of features to measure the similarity between different two-channel descriptors $H_{c_1,c_2}^{(i)}$ and $H_{c_1,c_2}^{(j)}$ obtained from different samples respectively. It is shown as Eq. (22).

$$D_{c_1,c_2}^{i,j}\left(H_{c_1,c_2}^{(i)}, H_{c_1,c_2}^{(j)}\right) = \sum_{k=1}^{M}\left(W_{f_{c_1,c_2}}^{k} \times \left(\frac{\left|H_{c_1,c_2}^{(i,k)} - H_{c_1,c_2}^{(j,k)}\right|}{\left|H_{c_1,c_2}^{(i,k)} + \mu_t\right| + \left|H_{c_1,c_2}^{(j,k)} + \mu_q\right|}\right)\right) \tag{22}$$

$$\mu_t = \sum_{k=1}^{M} H_{c_1,c_2}^{(i,k)}/M \tag{23}$$

$$\mu_q = \sum_{k=1}^{M} H_{c_1,c_2}^{(j,k)}/M \tag{24}$$

where M is the dimensionality of feature vectors, i, j represent two different samples, $H_{c_1,c_2}^{(i,k)}$ and $H_{c_1,c_2}^{(j,k)}$ are the k-th feature value of $H_{c_1,c_2}^{(i)}$ and $H_{c_1,c_2}^{(j)}$, μ_t and μ_q are utilized to avoid the denominator to be zero. Then, we can compute multiple weight Canberra distances for multiple two-channel bar-shape descriptors, defined as $D = \{D_{c_1,c_2}\}$, $c_1, c_2 \in \{L, a, b, sobel\}$ and $c_1 \neq c_2$.

(B). Multiple optimal distance pairs

Inspired by the voting theory [45], the relative distances based on these features pairs are not easy to make mistakes simultaneously and the relative distance of the optimal feature pairwise is minimal. In other words, the joint distance based on the optimal feature pairwise has a positive effect on person Re-ID. Higher similarity of the optimal distance pairs means lower relative distance and usually obtains lower risk of misrecognition. Based on this idea, we propose a novel metric learning method based on multiple optimal distance pairs via comparison of relative distance.

To capture multiple optimal distance pairs, we firstly normalize the distances to the range [0, 1]. Then, we construct dissimilarity matrix (***DM***) with different metrics to capture the multiple pairs of relative distance via weight Canberra distance as follows:

$$\begin{cases} DM_{i,j} = e^{|D_i - D_j|} \\ DC_{i,j} = e^{|D_i + D_j|} \\ D_i, D_j \in D \end{cases} \tag{25}$$

where $DM_{i,j}$ is the value of dissimilarity matrix at i-row and j-col, $DC_{i,j}$ is the value of matrix combining two-distance pairs and D_i, $D_j \in D$ are two different weight Canberra distances obtained by Section 3.2(A). In our approach, we consider 4 channels and obtain 6 two-channel bar-shape structural features. Thus the ***DM*** and ***DC*** are symmetric matrices with 6×6. Finally, we fuse the n optimal pairs of distance with minimal dissimilarity and their credibility, defined as:

$$DM_{MODR} = \sum_{k=1}^{n} \frac{DC_{up}^{k}}{DM_{up}^{k}} \tag{26}$$

where $1/DM_{up}^{k}$ is defined as the credibility, DM_{up}^{k} is the *k-th* element of the sorted upper triangular matrix of DM with ascending, DC_{up}^{k} is the *k-th* element of the sorted upper triangular matrix of DC associated with DM.

In general, the overall process of our algorithm (***MMLBD***) which integrates the two separate steps of feature representation and distance metric, is shown as follows.

Fig. 6. Examples of person Re-ID on the i-LIDS MCTS dataset.

Algorithm: the proposed method of MMLBD
Input: the dataset $X \in R^{128 \times 48}$, T, G, P.
Output: the rank of matching rates.

Begin:
(1) Obtain the channels of $L^*a^*b^*$ and *Sobel* space.
(2) Obtain single-channel bar-shape structural encoding, $(BSL_\theta^c(A), BSC_\theta^c(A))$ by Eqs.7–8.
(3) Obtain two-channel bar-shape structural encoding, $(BSL_\theta^{i,j}(A), BSC_\theta^{i,j}(A))$ by Eqs. 10–11.
(4) Obtain the color difference weights (W_θ) by Eqs. 14–16.
(5) Extract bar-shape structural Weight Color Difference Histogram and obtain the features:
$H_{i,j} = [HL_{0^\circ}^{i,j}, HC_{0^\circ}^{i,j}, HL_{45^\circ}^{i,j}, HC_{45^\circ}^{i,j}, HL_{90^\circ}^{i,j}, HC_{90^\circ}^{i,j}, HL_{135^\circ}^{i,j}, HC_{135^\circ}^{i,j}]$, $i, j \in \{L, a, b, sobel\}$.
(6) Divide X into training set (X_t) and test set which is made of a gallery set (X_g) and a probe set (X_p): $X_t(t), t = \{1, 2, \ldots, T\}$, $X_g(g), g = \{1, 2, \ldots, G\}$ and $X_p(p), p = \{1, 2, \ldots, P\}$.
(7) Learn the weight of feature $W_{f_{c_1,c_2}}$ by Eqs. 19–21.
(8) Obtain weight Canberra distance D_{c_1,c_2} by Eqs. 22–24.
(9) Construct dissimilarity matrix (DM) and two-distance pairs combining matrix (DC) by Eq. 25.
(10) Join distance metric via optimal distance pairs to compute the final distance D_{MODP} by Eq. 26.
(11) Obtain the matching rates with the theory of nearest neighbor.
End

3.3. Complexity analysis

In our approach, we consider the hybrid encoding strategy with multiple orientations and binary interaction mechanism, integrating multi-channel features. Compared with the original local binary pattern, the dimensionality of feature vector is reduced from $V^{8 \times n}$ to $\binom{n}{2} \times m \times V^{2\times 2}$, where n is the number of channels, m is the number of orientations and V is the maximum value of adjacent points with encoding or granularity. To be precise, the value of V for local binary encoding with 2-channels and multiple orientations is 2. Therefore, the dimensionality of $HL_\theta^{c_1,c_2}$ is $2^{2 \times 2} = 4$. Besides, for the color granularity encoding, the dimensionality of $HC_\theta^{c_1,c_2}$ is $k_{c_1} \times k_{c_2} \in \{16, 36\}$. Furthermore, the value of m is 4 and we can obtain that the dimensionality of $H_{i,j}$ is $\sum_\theta Dim(HL_\theta^{c_1,c_2}) + \sum_\theta Dim(HC_\theta^{c_1,c_2})$. Meanwhile, we take advantage of ***overlapping slide window*** with size of 16 × 16 to capture contrast-normalized local histogram and apply ***Max-pooling*** operator to fuse these local histograms on the horizontal direction. The step of slide window is denoted as 8. Thus, we can obtain that the final dimensionality of feature vector is $15 \times Dim(H_{c_1,c_2})$ from a person image with the size of 128×48. In addition, for our metric learning, we construct the optimal distance pairs to fuse multiple two-channel features and the complexity of time is $O_1(n^4) \times O_2(d)$ where d is the dimensionality of feature vector with subspace projection.

4. Experiments

4.1. Parameter and evaluation

We evaluate the proposed method of MMLBD on three-person Re-ID benchmark datasets, including i-LIDS Multiple-Camera Tracking Scenario (MCTS) [22], CAVIAR4REID [49] and Wide Area Re-Identification Dataset (WARD) [50]. Among them, the i-LIDS MCTS dataset evaluates the performance over variations of lighting illumination, obstacle, etc.; the CAVIAR4REID dataset focuses on the changes of resolution, pose as well as occlusions; the WARD dataset is concerned about a huge illumination variation apart from resolution and pose changes. Meanwhile, we resize each image in the datasets to 128×48 pixels to facilitate the evaluation with the common parameters of the descriptor. Besides, we chose the images of $\boldsymbol{p}$ persons (classes) to set up the training set, and the rest for testing. Each test set includes a gallery set and a probe set. The gallery set consists of one image (***single shot***) for each person and the remaining images are used as the probe set. This procedure is repeated 10 times.

We compare the proposed MMLBD with **LOMO+XQDA[9], MFA [24], kLFDA [24], eSDC [18], PRICoLBP [48], sub-mCT histogram [35], ELF [47], LBP [36] and HOG [46]** on three person Re-ID datasets. Besides, we extract the MBD descriptor from local overlapping slide windows with the size of 16×16 and the step is 8×8 on horizontal and vertical directions. Thus, for the horizontal direction, we can capture $\frac{48}{8} - 1 = 5$ local histograms. Furthermore, we can obtain a local maximal occurrence representation with max-pooling operator on horizontal direction. Besides, we can capture $\frac{128}{8} - 1 = 15$ local histograms on vertical direction. To be fair, we also set the same parameters of local path with the local descriptors of HOG, LBP, sub-mCT histogram and PRICoLBP. It is worth noting that we capture the LOMO descriptor from local path with the size of 10×10 and step of 5×5. It is basically optimal parameter setting, citing the original published paper. For the other method, we also ensure the optimal parameters based on the published papers. And they pay more attention to the metric learning method.

We utilize the standard performance measurements to evaluate our proposed **MMLBD**, also known as Cumulative Matching Characteristic (CMC) curve and Synthetic Disambiguation /Reacquisition Rate [51]. The CMC curve represents the expectation of the probe image correct match at rank r against the $\boldsymbol{p}$ gallery images. And rank-1 matching rate is thus the correct matching recognition rate. However, the SD/RR curve measures the probability that any of the $\boldsymbol{m}$ best matches is correct. In practice, a high rank-1 matching rate is significant, meanwhile, the top $\boldsymbol{r}$ ranked matching rates with a small value are also critical because the top matched images will

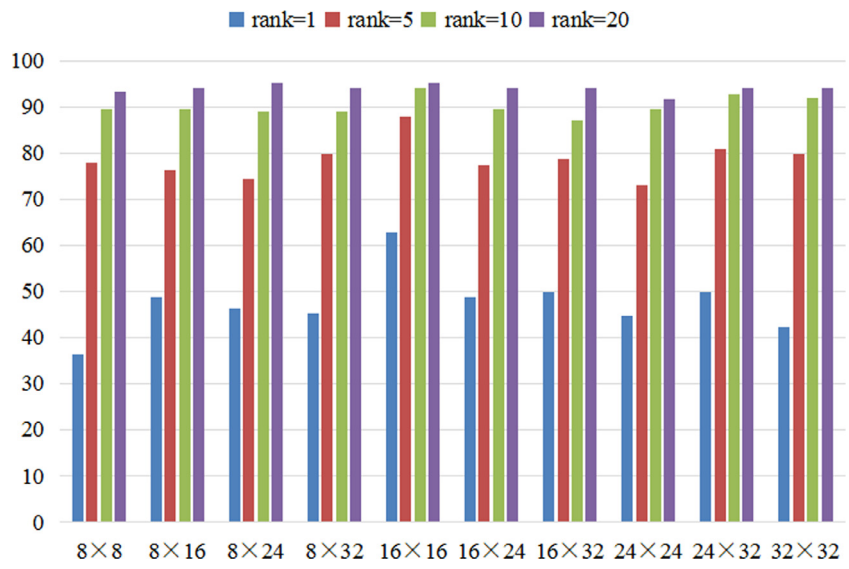
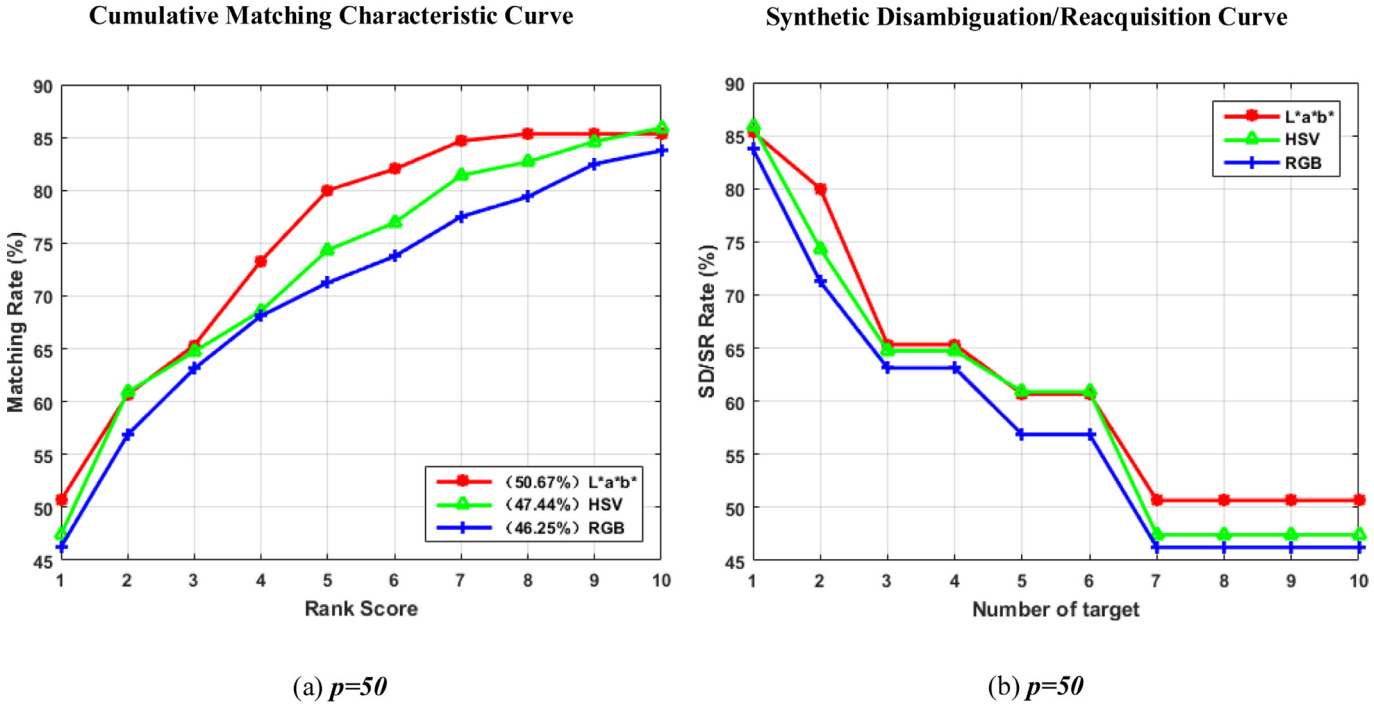


Fig. 7. The rank 1, 5, 10, 20 matching rates (%) with different cell scales on the i-LIDS MCTS dataset.



(a) ***p=50*** (b) ***p=50***

Fig. 8. The rank-1 accuracies, CMC and SD/RR curves of our approach with different color spaces (***RGB***, ***HSV*** and ***L*a*b****) on the i-LIDS MCTS dataset.

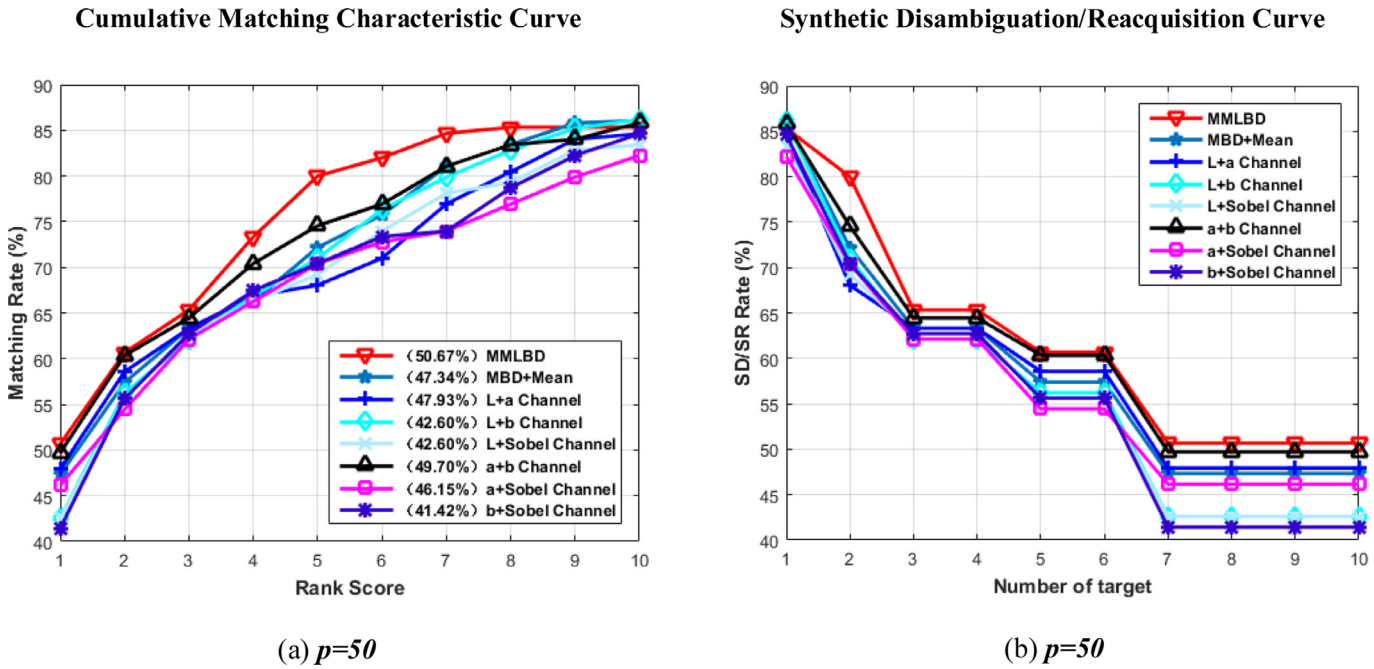
normally be verified by a human operator. The detailed results of experiments on person *Re*-ID are reported and analyzed below.

4.2. Experiments on i-LIDS multiple-camera tracking scenario (MCTS) dataset

In the i-LIDS MCTS dataset, in which the images are captured at an airport arrival hall during a busy period in a multi-camera CCTV network, there are a total of 119 persons and 476 images, as shown in Fig. 6. Firstly, experiments on i-LIDS MCTS dataset are carried out to evaluate the performance of the proposed method over the variations of lighting illumination, obstacle, and etc. Then we verified different impacts on performance over variations of cell scales, color spaces and two channel features respectively. Finally, we validated the performance of different feature descriptor and metric learning strategy.

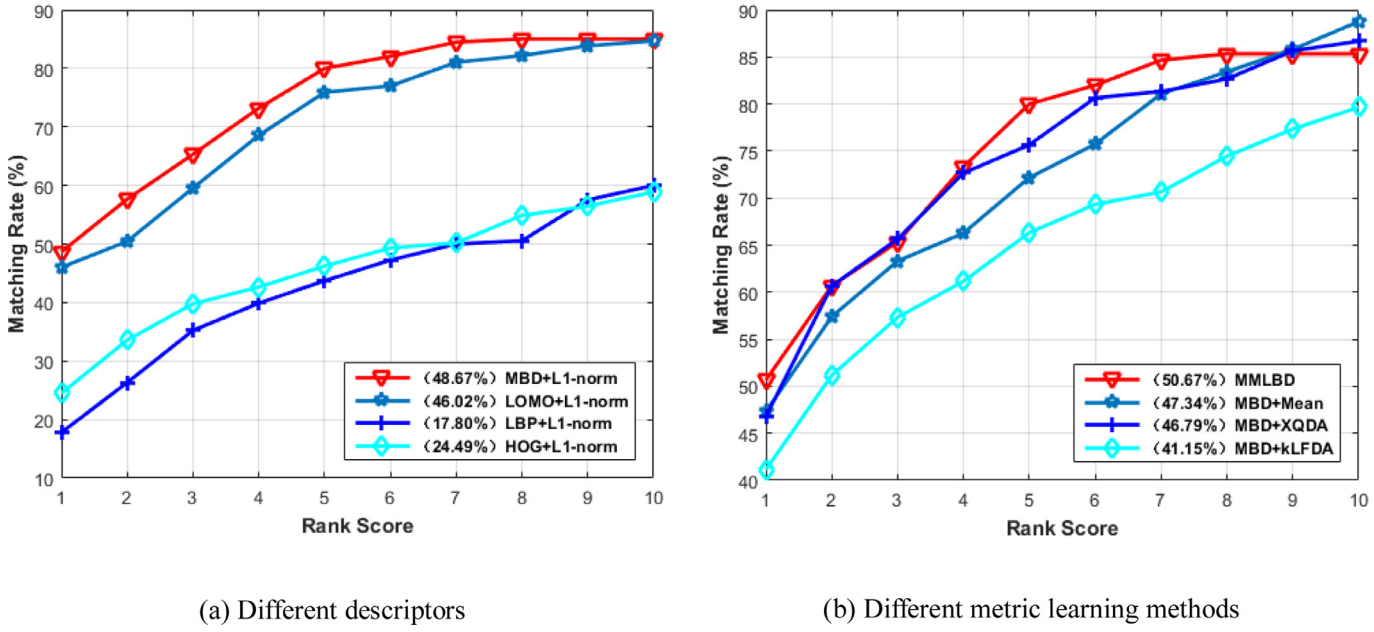
(A). Impact of different cell scales

In this section, we present the details and results of comparative experiments and further analyze the effects on performance under different cell scales, defined as the overlapping sliding windows by varying widths ($w = 8,\ 16,\ 24,\ 32$) and heights ($h = 8,\ 16,\ 24,\ 32$). The matching rates at rank-1, 5, 10, 20 with different scales of cell are reported in Fig. 7. It can be seen that the matching rates with the cell of 16×16 are higher than other cell scales at rank-1, 5, 10, 20 and it can achieve the best performance compared with other cells. Taking into account the dimensionality and matching rates as a whole, we empirically chose the size of cell as 16×16 in our experiments to ensure lower



(a) $p=50$ (b) $p=50$

Fig. 9. The rank-1 accuracies, CMC and SD/RR curves of our approach and different bar-shape structural descriptors on the i-LIDS MCTS dataset.



(a) Different descriptors (b) Different metric learning methods

Fig. 10. The rank-1 accuracies, CMC curves of our approach, different descriptors and metric learning methods on the i-LIDS MCTS dataset.

dimensionality of feature vectors and the overlapping of different cells.

In addition, we will testify the performance of the proposed approach of MMLBD in different color spaces of ***RGB***, ***HSV*** and ***L*a*b****, and the details are presented in the following section.

(B). Impact of different color spaces

The performance of our approach is evaluated on different color spaces ***RGB***, ***HSV*** and ***L*a*b****, and the results are shown in Fig. 8. In general, our proposed MMLBD demonstrates relatively robust performance in different color spaces, particularly in the color space of ***L*a*b****. We can achieve the best matching rate of 50.67% and a 3.23% performance gain can be obtained for the rank-1 accuracy with $\boldsymbol{p}=50$ since the ***L*a*b**** color space is a kind of color systems based on human visual physiological characteristics. In terms of person Re-ID, the color space of ***L*a*b**** is more suitable to reduce intra-class variations when comparing with the color spaces of ***RGB*** and ***HSV***. Therefore, we apply our presented method of MMLBD in the ***L*a*b**** color space for person Re-ID. To further demonstrate the effectiveness of the proposed algorithm of **MMLBD**, evaluation of different two-channel features is conducted and the results will be reported in the following section.

(C). Comparative experiments under different two-channel features

Comparative experiments under different two-channel features are presented. Fig. 9 reports the CMC curves and SD/RR curves and we can see that the proposed approach based on optimal distance pairs strategy can do better than other two-channel features, outperforming the second best one ($\mathbf{a}+\mathbf{b}$ channel) by 1.95%. It indi-

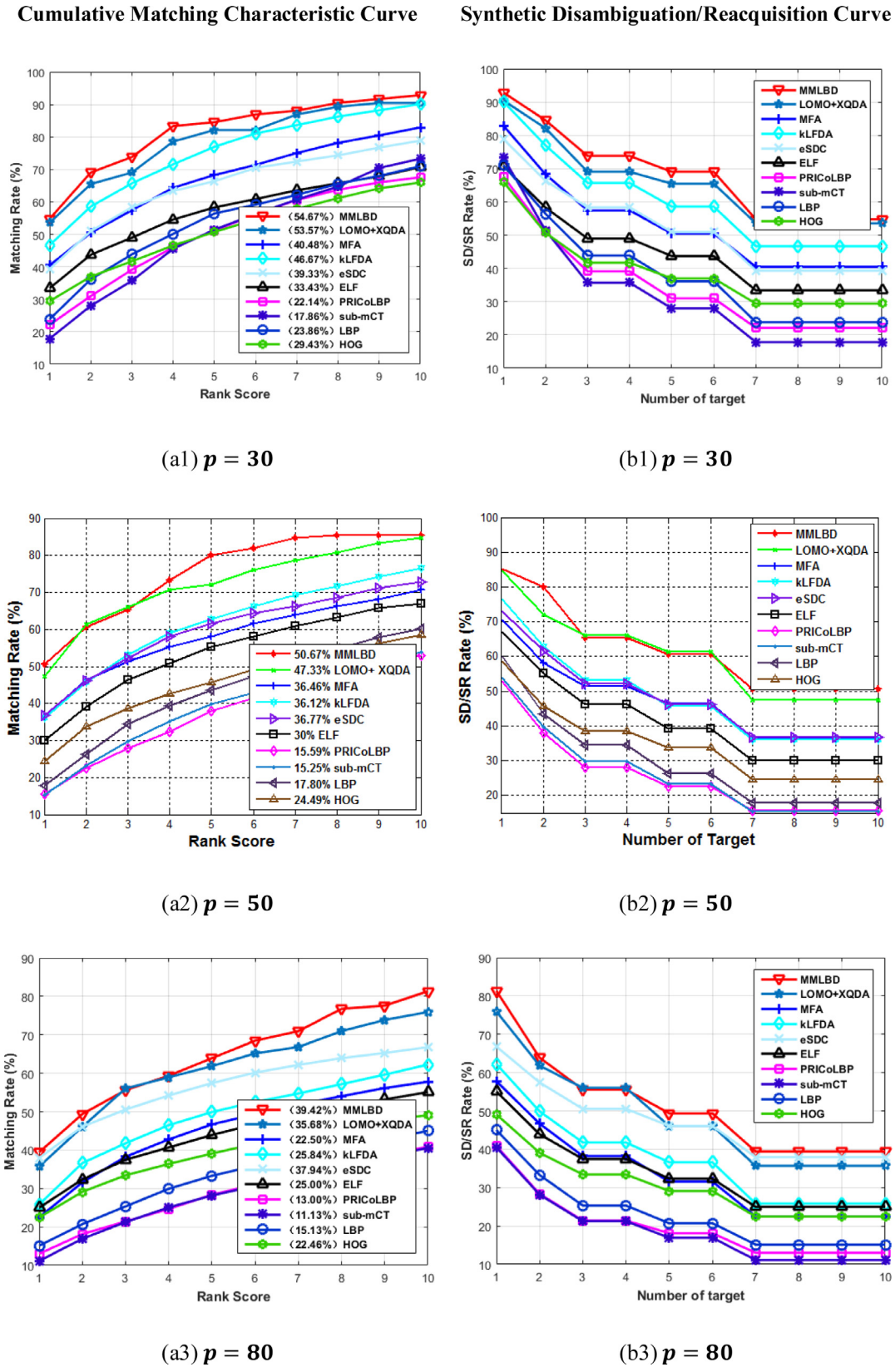


Fig. 11. The rank-1 accuracies, CMC and SD/RR curves on the i-LIDS MCTS dataset.

Fig. 12. Examples of person re-identification on the CAVIAR4REID dataset.

Table 2
The rank 1, 5, 10, 20 matching rates (%) with our approach, LOMO+XQDA, MFA, kLFDA, eSDC, PRICoLBP, sub-mCT histogram, ELF, LBP and HOG on the i-LIDS MCTS dataset.

Methods	$p = 30$				$p = 50$				$p = 80$			
	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 5$	$r = 10$	$r = 10$	$r = 20$	$r = 1$	$r = 5$	$r = 10$	$r = 20$
MMLBD	**54.76**	**84.52**	**92.86**	**95.24**	**50.67**	**80**	**85.33**	**94.67**	**39.42**	**63.90**	**81.34**	**87.24**
LOMO+ XQDA	53.57	82.14	90.48	97.61	47.33	72	84.67	93.33	35.68	61.83	75.93	87.13
MFA	40.48	68.27	82.92	96.07	36.46	58.16	70.65	86.22	22.50	46.68	57.78	72.58
kLFDA	46.67	76.96	90.12	97.62	36.12	62.76	76.46	90.71	25.84	49.90	62.23	75.88
eSDC	39.33	66.35	78.85	93.08	36.77	61.56	66.78	77.79	37.94	57.43	66.78	67.92
ELF	33.43	58.29	70.71	90.43	30	55.25	67.03	82.20	25	43.9	55.08	68.67
PRICoLBP	22.14	51	67.57	86.43	15.59	37.88	52.88	75.59	13	28.38	40.83	58.08
sub-mCT	17.86	51.43	73.29	90.71	15.25	39.75	53.8	73.90	11.13	28.17	40.50	59.08
LBP	23.86	56.43	71	88.29	17.80	43.56	60.08	76.69	15.13	33.29	45	60.63
HOG	29.43	50.71	66	88.29	24.49	45.76	58.56	73.73	22.46	39.08	49.08	61.58

cates that the proposed optimal distance pair strategy can capture more discrimination information via multiple optimal distance pairs which are relative distance obtained by dissimilarity matrix, and joint distance metric based on voting theory.

(D). Evaluation of different descriptor and metric learning

The proposed descriptor (**MBD**) and metric learning method (**MMLBD**) is compared with other descriptors (**LOMO, LBP, HOG**) and metric learning method **(MBD + Mean, MBD + XQDA, MBD + kLFDA**). For the different descriptors, the l_1-norm is utilized to measure the similarities between samples. Fig. 10 reports the CMC curves and the comparison shows that our proposed descriptor is obviously more robust than other descriptors and achieves the best matching rate of **50.67%** at rank-1 with $\boldsymbol{p = 50}$, over **3.23%** improvement than other descriptors, owing to the ensemble of color, texture and spatial structural information. Meanwhile, our proposed metric learning method also performs better than other metric learning methods, because of the consideration of discrimination information of multiple distance pairs. This indicates our proposed method of **MMLBD** designing a robust descriptor and optimal distance metric helps to reduce intra-class variations, so that the same person can be recognized at a higher rank.

(E). Comparison to the state of the art methods

In this section, all images (consisting of $\boldsymbol{p} = 30,\ 50,\ 80$ persons) are chosen to test the performance of the proposed **MMLBD**, compared with that of **LOMO + XQDA, MFA, kLFDA, eSDC, PRICoLBP, sub-mCT histogram, ELF, LBP and HOG** . The matching rates are shown in Fig. 11(a1, a2, a3) and Table 2. Compared with other approaches, our method achieves rank-1 matching rates of **54.76%, 50.67%, 39.42%** with $\boldsymbol{p} = 30,\ 50,\ 80$, and outperforms the best result obtained by **LOMO + XQDA**, which can only achieve a rank-1 matching rates **of 53.57%, 47.33%, 35.68%**, respectively. Furthermore, the advantage of our approach is obviously demonstrated at rank-5, 10, 20 with $\boldsymbol{p} = 30,\ 50,\ 80$. Besides, from the SD/RR curves in Fig. 11(b1, b2, b3), the performance of our method is also superior to that of others. The better performance of the proposed method indicates the following conclusions: (1) the proposed descriptor of **MBD** considers bar-shape structures with multi-orientation, and combines the spatial relation between points and their adjacent points. As a result, it can accurately capture more robust structural features underlying the pedestrian images. (2) The proposed descriptor of **MBD** integrates color difference information and eliminates the contrast between different blocks via overlapping strategy, which can improve the adaptability over the variance of illumination and shadowing. (3) The proposed descriptor of **MBD** capturing features from multiple scales is more robust to the changes of pose and camera views.

4.3. Experiments on CAVIAR4REID dataset

The CAVIAR4REID dataset which is extracted from the well-known CAVIAR dataset has been widely used to evaluate the performance of person Re-ID with resolution changes, light conditions, occlusions and pose changes. There are 72 pedestrians and 1220 images which consist of 50 persons captured from two different cameras in an indoor shopping center in Lisbon and 22 persons captured from only one camera and normalized to different sizes varying from 17×39 pixels to 72×144 pixels. Example images are shown in Fig. 12.

In this experiment, we chose all images consisting of $\boldsymbol{p} = 36,\ 50$ persons from this dataset and compare the rank-1, 2, 5, 10 matching rates of the proposed MMLBD with several state-of-the-art and correlative approaches (**LOMO+XQDA, MFA, kLFDA, eSDC, PRICoLBP, sub-mCT histogram, ELF, LBP and HOG**) in Table 3. It can be seen that our approach has an obvious advantage at rank-1, achieving the matching rates of **44.33%** and **35.26%** with $\boldsymbol{p} = 36,\ 50$, which are much higher than the best results of **34.92%**

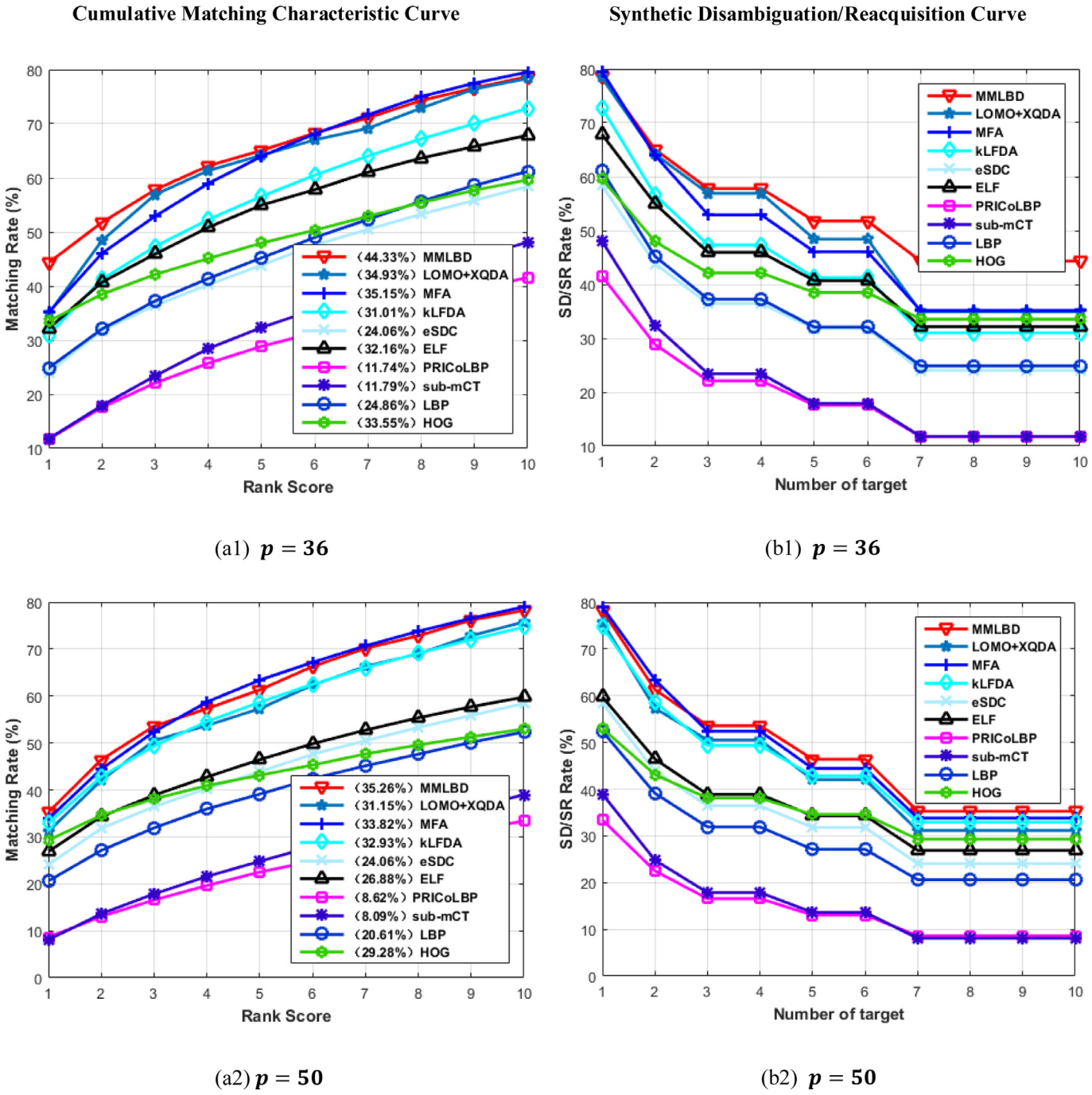


Fig. 13. The rank-1 accuracies, CMC and SD/RR curves on the CAVIAR4REID dataset.

Table 3
The rank 1, 2, 5, 10 matching rates (%) with our approach, LOMO+XQDA, MFA, kLFDA, eSDC, PRICoLBP, sub-mCT histogram, ELF, LBP and HOG on the CAVIAR4REID dataset.

Methods	$p = 36$				$p = 50$			
	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 1$	$r = 5$	$r = 10$	$r = 20$
MMLBD	**44.33**	**65.07**	**78.72**	**92.02**	**35.26**	**64.28**	**78.21**	**93.08**
LOMO+ XQDA	34.92	64.18	78.36	94.50	31.15	57.31	75.77	90.90
MFA	35.15	64.18	**79.56**	94.56	33.82	57.31	78.97	**93.15**
kLFDA	31.01	63.98	72.81	92.29	32.93	63.38	74.67	91.91
eSDC	24.06	47.60	58.41	78.92	23.30	42.25	53.50	69.21
ELF	32.16	54.96	67.87	85.12	26.88	46.44	59.74	59.35
PRICoLBP	11.74	28.87	41.58	64.82	8.62	22.50	33.39	50.74
sub-mCT	11.79	32.36	48.10	71.88	8.09	24.79	38.94	76.26
LBP	24.88	45.23	61.19	82.16	20.61	39.10	52.35	69.09
HOG	33.55	47.96	59.63	78.42	29.28	43.11	52.99	66.89

Fig. 14. Examples of person Re-ID on the WARD dataset: (A) the examples person acquired by camera A; (B) the examples person acquired by camera B; (C) the examples person acquired by camera C.

Table 4
The rank 1, 5, 10 matching rates (%) with our approach, LOMO+XQDA, MFA, kLFDA, eSDC, PRICoLBP, sub-mCT histogram, ELF, LBP and HOG on the WARD dataset.

Methods	Camera 1–2			Camera 1–3			Camera 2–3		
	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$
MMLBD	**67.89**	**89.98**	**95.98**	**63.67**	**85.95**	**94.10**	**76.89**	**92.35**	**95.93**
LOMO+ XQDA	58.65	84.93	94.35	48.38	77.89	89.36	61.26	86.70	93.93
MFA	47.84	66.63	79.38	42.57	61.42	73.69	53.03	71.94	79.81
kLFDA	45.15	53.44	57.72	42.95	51.78	57.52	49.54	64.27	72.16
eSDC	45.27	52.52	54.93	40.90	49.30	50.93	47.03	53.44	56.31
ELF	52.50	77.21	86.95	47.49	66.20	77.68	73.35	86.49	91.48
PRICoLBP	14.45	31.83	43.68	14.73	34.06	46.10	25.49	45.33	56.28
sub-mCT	24.99	48.12	62.29	23.16	42.70	53.95	28.13	52.58	65.30
LBP	33.34	59.21	71.17	34.12	54.53	67.71	58.56	78.63	86.59
HOG	43.76	59.75	67.62	48.13	65.36	74.11	51.13	67.39	75.77

and **33.82%** obtained by the approach of **MFA**. In few cases, such as rank = 10, 20, the performance of our approach is slightly lower that of **MFA**. However, in practice, the top $\boldsymbol{r}(\boldsymbol{r} \leq 10)$ ranked matching rates are more critical because the top matched images will normally be verified by a human operator. Obviously, in the case of rank $r(r \leq 10)$, it can be observed that **MMLBD** consistently performs the best. Therefore, in general, the proposed **MMLBD** outperforms existing state of the art methods. From Fig. 13, we can see that the performance of **MMLBD** is superior to that of others in most cases. In other words, the proposed **MMLBD** is robust to light conditions, occlusion, resolution and pose changes, etc. The reason is that the novel weight of features helps to capture more sufficient discrimination information and the proposed metric learning strategy selects the optimal distance pairs, which is effective to represent the appearance of person with occlusion, pose changes, etc.

4.4. Experiments on wide area re-identification dataset (WARD) dataset

The WARD dataset consists of 70 different pedestrians and 4786 images which are acquired by three non-overlapping cameras in a real surveillance scenario. This dataset is interesting owing to having a huge illumination variation apart from resolution and pose changes, as shown in Fig. 14. The dataset has three different cameras and we conduct the experiments for all different camera pairs, denoted as camera pairs 1–2, 1–3 and 2–3. The 70 pedestrians in this dataset are divided into training set and test set containing 60 and 10 persons respectively.

In this experiment, we compare the proposed **MMLBD** with other methods (**LOMO+XQDA, MFA, kLFDA, eSDC, PRICoLBP, sub-mCT histogram, ELF, LBP and HOG**). Experimental results are shown in Fig. 15 and Table 4. It can be seen that the **MMLBD** obviously outperforms other methods. For camera pairs 1–2, 1–3 and 2–3, the best rank-1 matching rates of **MMLBD** with $\boldsymbol{p} = 60$ are **67.89%, 63.67% and 76.89%,** increased by **11.24%, 15.29%** and **15.63%,** respectively. Meanwhile, the SD/RR curve also shows that the performance of **MMLBD** is superior to that of others.

From the analysis of all the reported results, we can conclude that, in general, our method has superior performance than state-of-the-art approaches. This is supported by the fact that we achieve the best overall performance in CMC and SD/RR curves for all the three considered datasets.

5. Conclusions

In this paper, we design a novel person Re-ID algorithm (**MMLBD**) inspired by relative distance comparison and multi-channel visual features. The proposed method is proved to demonstrate an outstanding performance for the representation of person's appearance. In our approach, we capture the intrinsic structure information hidden in different person images through multiple bar-shape descriptor that make full use of spatial correlation between center points and their neighbors. Also, we propose a new color difference weight for fusion of color information and apply an overlapping strategy to reduce the local contrast problem in images. Considering the human visual mechanism, we granulate and encode the multi-channel color information into more

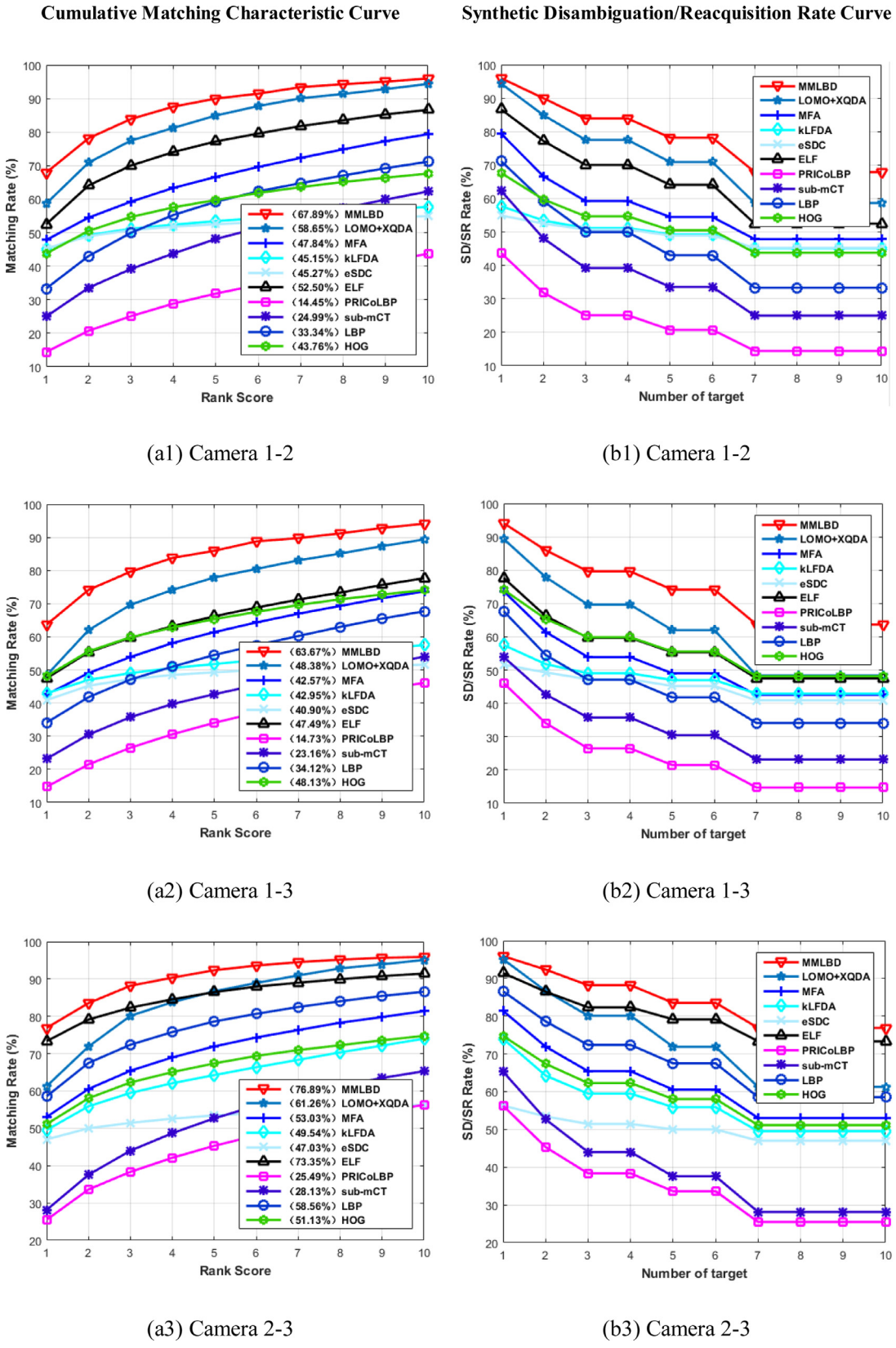


Fig. 15. The rank-1 accuracies, CMC and SD/RR curves on the WARD dataset.

rough blocks with the color space ***L*a*b**** and ***Sobel*** channel, instead of gray-values. Thus, it can do better to represent the appearance of person with the changes of illumination, rotation, translation and perspective for person Re-ID. Meanwhile, we present a novel multiple metric learning method based on the similarity and dissimilarity of different samples. For bar-shape structural descriptors with multiple orientations, we make use of minimum relative distance and dissimilarity matrix to learn the most suitable combination for final distance metric and obtain the weight based on the comparison of relative distance. Thus, the influence of the noise or outliers can be diminished. On the whole, the proposed **MMLBD** is simple but effective. Finally, the experimental results have demonstrated that the proposed **MMLBD** outperforms the **LOMO+XQDA, MFA, kLFDA, eSDC, ELF, PRICoLBP, sub-mCT, LBP** and **HOG**.

Acknowledgements

The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions. This work is partially supported by China National Natural Science Foundation under grant No. 61673299, 61203247, 61573259 and 61573255. It is also partially supported by Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University) under grant No. MJUKF201721. It is also supported by the Fundamental Research Funds for the Central Universities (Grant No. 2013KJ010). It is also partially supported by the Open Project Program of Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education under grant No. 30920130122005. It is also partially supported by the program of Further Accelerating the Development of Chinese Medicine Three Year Action of Shanghai grant No. ZY3-CCCX-3-6002, and the Research Grant of the Hong Kong Polytechnic University (Grant No. G-YM53).

References

[1] S. Gong, M. Cristani, S. Yan, C.C. Loy, Person Re-Identification, vol. 1, Springer, 2014.
[2] L. Zheng, Y. Yang, and A. G. Hauptmann. Person re-identification: past, present and future. arXiv preprint arXiv:1610.02984, 2016.
[3] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, S.Z. Li, Salient color names for person re-identification, in: European Conference on Computer Vision, Springer, 2014, pp. 536–551.
[4] A.I. Awad, M. Hassaballah, Image Feature Detectors and Descriptors, vol. 630, Springer, 2016.
[5] C. Zhao, D. Miao, Z. Lai, C. Gao, C. Liu, J. Yang, Two-dimensional color uncorrelated discriminant analysis for face recognition, Neurocomputing 113 (2013) 251–261.
[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, Person re-identification by symmetry-driven accumulation of local features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2360–2367.
[7] H. Jegou, F. Perronnin, M. Douze, J. Sánchez, P. Perez, C. Schmid, Aggregating local image descriptors into compact codes, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1704–1716.
[8] L. Ma, H. Liu, L. Hu, C. Wang, and Q. Sun. Orientation driven bag of appearances for person re-identification. arXiv preprint arXiv:1605.02464, 2016.
[9] S. Liao, Y. Hu, X. Zhu, S.Z. Li, Person re-identification by local maximal occurrence representation and metric learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2197–2206.
[10] T. Matsukawa, T. Okabe, E. Suzuki, Y. Sato, Hierarchical Gaussian descriptor for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1363–1372.
[11] Y. Yan, B. Ni, Z. Song, C. Ma, Y. Yan, X. Yang, Person re-identification via recurrent feature aggregation, in: European Conference on Computer Vision, 2016, pp. 701–716.
[12] F. Shen, C. Shen, Q. Shi, A. van den Hengel, Z. Tang, H.T. Shen, Hashing on nonlinear manifolds, IEEE Trans. Image Process. 24 (6) (2015) 1839–1851.
[13] S. Ding, L. Lin, G. Wang, H. Chao, Deep feature learning with relative distance comparison for person re-identification, Pattern Recognit. 48 (10) (2015) 2993–3003.
[14] D. Cheng, Y. Gong, S. Zhou, J. Wang, N. Zheng, Person re-identification by multi-channel parts-based cnn with improved triplet loss function, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1335–1344.
[15] J. Wang, Z. Wang, C. Gao, N. Sang, R. Huang, Deeplist: learning deep features with adaptive listwise constraint for person re-identification, IEEE Trans. Circuits Syst. Video Technol. 27 (3) (2016) 1–12.
[16] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.
[17] P.-E. Forssén, Maximally stable colour regions for recognition and matching, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
[18] R. Zhao, W. Ouyang, X. Wang, Unsupervised salience learning for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3586–3593.
[19] S. Iodice, A. Petrosino, Salient feature based graph matching for person re-identification, Pattern Recognit. 48 (4) (2015) 1074–1085.
[20] Y. Shen, W. Lin, J. Yan, M. Xu, J. Wu, J. Wang, Person re-identification with correspondence structure learning, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3200–3208.
[21] W. Hu, M. Hu, X. Zhou, T. Tan, J. Lou, S. Maybank, Principal axis-based correspondence between multiple cameras for people tracking, IEEE Trans. Pattern Anal. Mach. Intell. 28 (4) (2006) 663–671.
[22] W.-S. Zheng, S. Gong, T. Xiang, Reidentification by relative distance comparison, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (2013) 653–668.
[23] S. Pedagadi, J. Orwell, S. Velastin, B. Boghossian, Local fisher discriminant analysis for pedestrian re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3318–3325.
[24] F. Xiong, M. Gou, O. Camps, M. Sznaier, Person reidentification using kernel-based metric learning methods, in: European Conference on Computer Vision, 2014, pp. 1–16.
[25] P.M. Roth, M. Hirzer, M. Koestinger, C. Beleznai, H. Bischof, Mahalanobis distance learning for person reidentification, in: Person Re-Identification, 2014, pp. 247–267.
[26] L. Ma, X. Yang, D. Tao, Person re-identification over camera networks using multi-task distance metric learning, IEEE Trans. Image Process. 23 (8) (2014) 3656–3670.
[27] G. Lisanti, I. Masi, A.D. Bagdanov, A. Del Bimbo, Person re-identification by iterative re-weighted sparse ranking, IEEE Trans. Pattern Anal. Mach. Intell. 37 (8) (2015) 1629–1642.
[28] N. Martinel, C. Micheloni, G.L. Foresti, Kernelized saliency-based person re-identification through multiple metric learning, IEEE Trans. Image Process. 24 (12) (2015) 5645–5658.
[29] E. Ahmed, M. Jones, T.K. Marks, An improved deep learning architecture for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3908–3916.
[30] Z. Lai, W.K. Wong, Y. Xu, C. Zhao, M. Sun, Sparse alignment for robust tensor learning, IEEE Trans. Neural Netw. Learn. Syst. 25 (10) (2014) 1779–1792.
[31] R. Zhao, W. Ouyang, X. Wang, Person re-identification by saliency learning, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2) (2016) 356–370.
[32] D. Tao, Y. Guo, M. Song, Y. Li, Z. Yu, Y.Y. Tang, Person re-identification by dual-regularized kiss metric learning, IEEE Trans. Image Process. 25 (6) (2016) 2726–2738.
[33] W.-S. Zheng, S. Gong, T. Xiang, Towards open-world person re-identification by one-shot group-based verification, IEEE Trans. Pattern Anal. Mach. Intell. 38 (3) (2016) 591–606.
[34] J. Wu, J.M. Rehg, Centrist: a visual descriptor for scene categorization, IEEE Trans. Pattern Anal. Mach. Intell. 33 (8) (2011) 1489–1501.
[35] Y. Xiao, J. Wu, J. Yuan, mcentrist: a multi-channel feature generation mechanism for scene categorization, IEEE Trans. Image Process. 23 (2) (2014) 823–836.
[36] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[37] Z. Guo, L. Zhang, D. Zhang, A completed modeling of local binary pattern operator for texture classification, IEEE Trans. Image Process. 19 (6) (2010) 1657–1663.
[38] X. Fu, W. Wei, Centralized binary patterns embedded with image Euclidean distance for facial expression recognition, in: Fourth IEEE International Conference on Natural Computation, 2008, pp. 115–119.
[39] B. Zhang, Y. Gao, S. Zhao, J. Liu, Local derivative pattern versus local binary pattern: face recognition with highorder local pattern descriptor, IEEE Trans. Image Process. 19 (2) (2010) 533–544.
[40] W.-H. Liao, Region description using extended local ternary patterns, in: IEEE International Conference on Pattern Recognition, 2010, pp. 1003–1006.
[41] S.K. Vipparthi, S.K. Nagar, Color directional local quinary patterns for content based indexing and retrieval, Human Centric Comput. Inf. Sci. 4 (1) (2014) 1.
[42] W.-H. Liao, C.-Y. Liu, M.-C. Lin, Feature description using center-symmetric extended local ternary patterns, in: IEEE International Symposium on Multimedia, 2014, pp. 94–97.
[43] F. Shen, X. Zhou, Y. Yang, J. Song, H.T. Shen, D. Tao, A fast optimization method for general binary code learning, IEEE Trans. Image Process. 25 (12) (2016) 5610–5621.
[44] G.-H. Liu, J.-Y. Yang, Content-based image retrieval using color difference histogram, Pattern Recognit. 46 (1) (2013) 188–198.
[45] R.E. Schapire, Y. Freund, P. Bartlett, W.S. Lee, Boosting the margin: a new explanation for the effectiveness of voting methods, Ann. Stat. 26 (5) (1998) 1651–1686.

[46] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.
[47] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: European Conference on Computer Vision, Springer, 2008, pp. 262–275.
[48] X. Qi, R. Xiao, C.-G. Li, Y. Qiao, J. Guo, X. Tang, Pairwise rotation invariant co-occurrence local binary pattern, IEEE Trans. Pattern Anal. Mach. Intell. 36 (11) (2014) 2199–2213.
[49] T. Avraham, I. Gurvich, M. Lindenbaum, S. Markovitch, Learning implicit transfer for person re-identification, in: European Conference on Computer Vision, Springer, 2012, pp. 381–390.
[50] N. Martinel, A. Das, C. Micheloni, A.K. Roy-Chowdhury, Re-identification in the function space of feature warps, IEEE Trans. Pattern Anal. Mach. Intell. 37 (8) (2015) 1656–1669.
[51] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recognition, reacquisition, and tracking, Processing of IEEE International Workshop on Performance Evaluation for Tracking and Surveillance, vol. 3, 2007.

Cairong Zhao is currently an associate professor at Tongji University. He received the PhD degree from Nanjing University of Science and Technology, M.S. degree from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, and B.S. degree from Jilin University, in 2011, 2006 and 2003, respectively. He has published more than 20 scientific papers in pattern recognition, computer vision and related areas. His research interests include computer vision, pattern recognition and visual surveillance.

Xuekuan Wang is currently a master candidate in College of Electronics and Information Engineering, Tongji University. His research interests include computer vision, pattern recognition and machine learning, in particular, focusing on person re-identification for visual surveillance.

W. K. Wong received his Ph.D. degree from The Hong Kong Polytechnic University. Currently, he is with Institute of Textiles & Clothing, The Hong Kong Polytechnic University, Hong Kong and The Hong Kong Polytechnic University Shenzhen Research Institute. He has published more than fifty scientific articles in refereed journals, including IEEE Transactions on Neural Networks and Learning Systems, Pattern Recognition, International Journal of Production Economics, European Journal of Operational Research, International Journal of Production Research, Computers in Industry, IEEE Transactions on Systems, Man, and Cybernetics, among others. His recent research interests include artificial intelligence, pattern recognition, and optimization of manufacturing scheduling, planning and control.

Weishi Zheng received the Ph.D. degree in Applied Mathematics from Sun Yat-Sen University, in 2008. He is now a Professor at Sun Yat-sen University. He had been a Postdoctoral Researcher on the EU FP7 SAMURAI Project at Queen Mary University of London and an Associate Professor at Sun Yat-sen University after that. He has now published more than 80 papers, including more than 50 publications in main journals (TPAMI, TNN, TIP, TSMC-B, PR) and top conferences (ICCV, CVPR, IJCAI, AAAI). He has joined the organization of four tutorial presentations in ACCV 2012, ICPR 2012, ICCV 2013 and CVPR 2015 along with other colleagues. His research interests include person/object association and activity understanding in visual surveillance. He has joined Microsoft Research Asia Young Faculty Visiting Program. He is a Recipient of Excellent Young Scientists Fund of the National Natural Science Foundation Of China, and a recipient of Royal Society-Newton Advanced Fellowship.

Jian Yang received the BS degree in mathematics from the Xuzhou Normal University in 1995. He received the MS degree in applied mathematics from the Changsha Railway University in 1998 and the PhD degree from the Nanjing University of Science and Technology (NUST), on the subject of pattern recognition and intelligence systems in 2002. In 2003, he was a postdoctoral researcher at the University of Zaragoza. From 2004 to 2006, he was a Postdoctoral Fellow at Biometrics Centre of Hong Kong Polytechnic University. From 2006 to 2007, he was a Postdoctoral Fellow at Department of Computer Science of New Jersey Institute of Technology. Now, he is a professor in the School of Computer Science and Technology of NUST. He is the author of more than 80 scientific papers in pattern recognition and computer vision. His journal papers have been cited more than 1600 times in the ISI Web of Science, and 2800 times in the Web of Scholar Google. His research interests include pattern recognition, computer vision and machine learning. Currently, he is an associate editor of Pattern Recognition Letters and IEEE Transactions on Neural Networks and learning systems.

Duoqian Miao is currently a full professor and vice dean of the school of Electronics and Information Engineering of Tongji University. He received his PhD in Pattern Recognition and Intelligent System at Institute of Automation, Chinese Academy of Sciences in 1997. He works for Department of Computer Science and Technology of Tongji University, Computer and Information Technology Teaching Experiment Center, and the Key Laboratory of "Embedded System and Service Computing" Ministry of Education. He has published over 180 scientific articles in international journals, books, and conferences. He is committee member of International Rough Sets Society, senior member of China Computer Federation (CCF), committee member of CCF Artificial Intelligence and Pattern Recognition, committee member of Chinese Association for Artificial Intelligence (CAAI), chair of CAAI Rough Set and Soft Computing Society and committee member of CCAI Machine Learning, committee member of Chinese Association of Automation(CAA) Intelligent Automation, committee member and chair of Shanghai Computer Society(SCA) Computing Theory and Artificial Intelligence. His current research interests include: Rough Sets, Granular Computing, Principal Curve, Web Intelligence, and Data Mining etc.