Pattern Recognition 79 (2018) 79–96

Contents lists available at ScienceDirect

Pattern Recognition

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

Maximal granularity structure and generalized multi-view discriminant analysis for person re-identification

Cairong Zhao[a,1,*], Xuekuan Wang[a,1], Duoqian Miao[a,*], Hanli Wang[a], Weishi Zheng[b], Yong Xu[c], David Zhang[d]

[a] Department of Computer Science and Technology, Tongji University, Shanghai, China
[b] School of Information Science and Technology, Sun Yat-sen University, Guangzhou, China
[c] Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, China
[d] Biometrics Research Centre, Department of Computing, Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Received 10 July 2017
Revised 13 January 2018
Accepted 28 January 2018
Available online 2 February 2018

Keywords:
Person re-identification
Maximal granularity structure descriptor
Generalized multi-view discriminant analysis
Representation consistency

ABSTRACT

This paper proposes a novel descriptor called Maximal Granularity Structure Descriptor (MGSD) for feature representation and an effective metric learning method called Generalized Multi-view Discriminant Analysis based on representation consistency (GMDA-RC) for person re-identification (Re-ID). The proposed descriptor of MGSD captures rich local structural information from overlapping macro-pixels in an image, analyzes the horizontal occurrence of multi-granularity and maximizes the occurrence to extract a robust representation for viewpoint changes. As a result, the proposed descriptor of MGSD can obtain rich person appearance whilst being robust against different condition changes. Besides, considering multi-view information, we present a new GMDA-RC for different views, inspired by the observation that different views share similar data structures. The proposed metric learning method of GMDA-RC seeks multiple discriminant common spaces for multiple views by jointly learning multiple view-specific linear transforms. Finally, we evaluate the proposed method of (MGSD+GMDA-RC) on three publicly available person Re-ID datasets: VIPeR, CUHK-01 and Wide Area Re-ID dataset (WARD). For the VIPeR and CUHK-01, the experimental results show that our method significantly outperforms the state-of-the-art methods, achieving the rank-1 matching rates of 67.09%, 70.61%, and the improvements of 17.41%, 5.34%, respectively. For the WARD, we consider different pairwise camera views (camera 1–2, camera 1–3, camera 2–3) and our method can achieve the rank-1 matching rates of 64.33%, 59.42%, 70.32%, increasing of 5.68%, 11.04%, 9.06% compared with the state-of-the-art methods, respectively.

1. Introduction

Person re-identification (Re-ID) is the task of recognizing pedestrians observed from non-overlapping camera views in a surveillance system and it a challenging problem because of big intra-class variations in illumination, pose, viewpoint and occlusion [1,2]. To address this problem, existing approaches mainly focus on developing effective feature representation methods which are robust against the view/pose/illumination/background changes [3–26], or learning a distance metric [6–8]. The main development of person Re-ID have been shown in Table 1.

For feature representation, lots of models extract the low-level critical visual features (i.e. color [3–5], texture [6,7], structure [8–10], etc.), and consider the rich information of pedestrian's appearance. However, these handcrafted features would been constrained by scenarios for person Re-ID. In practice, most of the appearance-based approaches would integrate multiple features, such as ensemble of localized features [11], local maximal occurrence representation [12], salient match [13,14], ensemble of invariant features [15], camera correlation aware feature augmentation [16], kernel-based features [17], etc. Meanwhile, the semantic features [18–21] are also important for person Re-ID. Better yet, deep learning is also a noteworthy method, exhibiting an excellent performance in learning representation of person Re-ID [20,22–26]. These handcrafted or learning based descriptors have made impressive improvements over feature extraction, and advanced the person Re-ID research. Unfortunately, it is still extremely difficult to extract a robust feature that effectively adapts to severe changes and misalignment across disjoint views.

Another aspect of person Re-ID considers to design an effective metric learning model trying to seek a new discriminative

* Corresponding authors.
E-mail addresses: zhaocairong@tongji.edu.cn (C. Zhao), wxktongji@163.com (X. Wang), dqmiao@tongji.edu.cn (D. Miao).
[1] The authors contribute equally to this work.

https://doi.org/10.1016/j.patcog.2018.01.033
0031-3203/

Table 1
The main development of person Re-ID.

Author	Year	Approaches	Remark
Gray et al. [11]	2008	ELF	Appearance
Kostinger et al. [51]	2012	KISSME	Metric
Igor et al. [3]	2013	Color Invariants	Appearance
Zhao et al. [14]	2013	Salience	Appearance
Zheng et al. [27]	2013	RDC	Metric
Pedagadi et al. [28]	2013	LFDA	Metric
Yang et al. [4]	2014	Salient Color Name	Appearance
Xiong et al. [29]	2014	Kernel	Metric
Liao et al. [12]	2015	LOMO+XQDA	Appearance/metric
Shi et al. [18]	2015	Transfer	Metric
Li et al. [19]	2015	Attribute	Appearance
Ahmed et al. [36]	2015	Deep Learning	Appearance/metric
Paisitkriangkrai et al. [40]	2015	Rank	Metric
Matsukawa et al. [5]	2016	GOG	Appearance
Xiao et al. [22]	2016	Deep Learning	Appearance
Zhang et al. [30]	2016	Null Space Learning	Metric
Tao et al. [32]	2016	DR-KISS	Metric
Zheng et al. [33]	2016	Transfer	Metric
Peng et al. [34]	2016	Transfer	Metric
Yang et al. [53]	2016	LSSL	Metric
Zhao et.al [54]	2016	MLAPG	Metric
Zhang et al. [8]	2017	Structured Matching	Appearance/metric
Chen [16]	2017	CRAFT	Appearance/metric
Wang et al. [26]	2017	Deep Learning	Appearance/metric
Zhao et al. [31]	2017	Salience	Metric

subspace for more good performance. Typical algorithms include relative distance comparison (RDC) [27], local fisher discriminant analysis (LFDA) [28], kernel-based metric [29], cross-view quadratic discriminant analysis (XQDA) [12], MLAPG [54], discriminative null space [30], saliency learning [31], dual-regularized KISS (DR-KISS) [32], transfer learning [33,34] and deep learning model [35–39,43], etc. In addition, many other kinds of methods try to address Re-ID by ranking methods [40,41]. Although these metric-based methods outperform the existing Re-ID benchmarks, they are nevertheless limited by some of classical problems, such as the inconsistent distributions of multiple views and small sample size (SSS) for model learning.

For the multi-view learning, it has been utilized to solve various computer visual tasks, such as image annotation [58], image retrieval [60] and so on. In this paper, we design a novel robust feature representation called maximal granularity structure descriptor (MGSD) and an efficient metric learning method called generalized multi-view discriminant analysis with representation consistency (GMDA-RC). More specifically, based on the information granularity theory [42] and human visual attention mechanism [43], we preprocess the original images with multiple scales and orientations, and obtain the multi-granularity feature maps. To uncover the intrinsic relationship of different features, we design a local crossing coding method to capture salient features from a biologically inspired feature (BIF) magnitude image obtained by multiple Gabor filters [44]. Then, we analyze the horizontal occurrence of the local features and take advantage of MAX operator [12] to capture a robust representation against viewpoint changes. Furthermore, to learn an effective and robust distance or similarity function, we propose a novel metric measuring method of GMDA-RC, which can learn a low dimensional consistent discriminant subspaces from multiple views [45–47,56–59] and we solve this problem as a classic generalized eigenvalue decomposition problem [48]. This processing is shown in Fig. 1.

1.1. Motivation

For person Re-ID, how to extract a robust feature and learning an optimal distance metric across camera views are important problems. Existing handcrafted or learning methods have been shown to be effective in improving the person Re-ID benchmarks, yet they have some drawbacks as follows:(1) The traditional descriptors could characterize the certainty of different features, but fail for the uncertainty. However, fuzziness and uncertainty embedding image is very important property for Re-ID from the viewpoint of human perception; (2) Most of the existing methods assume that the distributions of multiple camera views is consistent. However, this assumption is one-sided because the important attribute of each camera view is different in practice; (3) Most of metric learning method suffer from the small sample size (SSS) problem.

To address the above-mentioned problems, we propose a novel maximal granularity structure descriptor (GMSD) from the view of feature extraction and a generalized multi-view discriminant analysis with representation consistency (GMDA-RC) from the view of metric learning for person Re-ID.

1.2. Contribution

The main contributions of our work are summarized as the following three points:

1. We introduce a new maximal granularity structure descriptor (MGSD) which extracts local salient features to describe pedestrian's appearance. To make a stable representation against viewpoint changes, we exploit a novel strategy of local maximal crossing coding to combine colors, textures and colors differences in color and orientation blocks, which not only considers the local features, but also the spatial relationships from different scales and texture orientations.
2. We propose a novel similarity measure to seek a low dimensional consistent discriminant subspace by generalized multi-view discriminant analysis with representation consistency (GMDA-RC), solved by the generalized eigenvalue decomposition. For the representation consistency, we minimize the error of construction form view-1 to view-2 and address it by least square method.
3. Benefitting from the consideration of feature representation and metric learning, the proposed method is shown to be effective

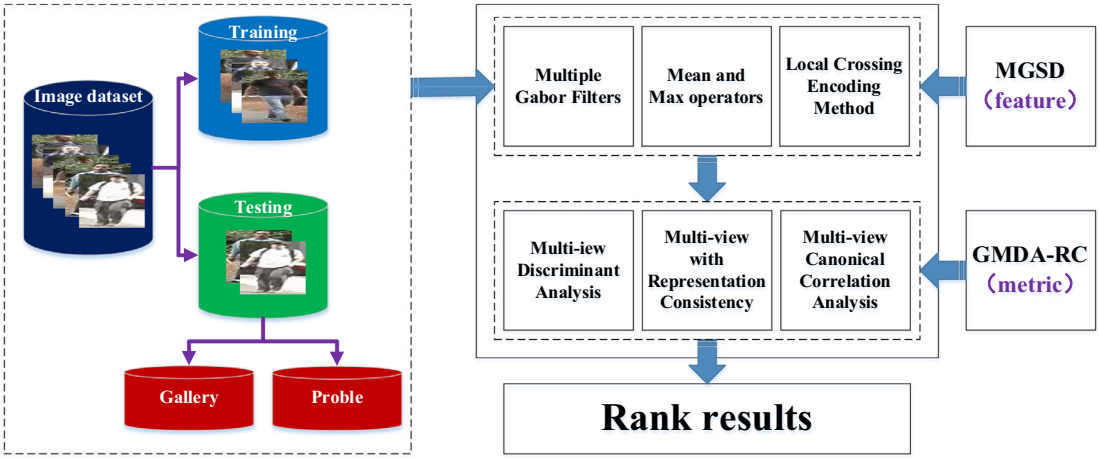


Fig. 1. The processing of person Re-ID based on the proposed method of MGSD+GMDA-RC: a maximal granularity structure descriptor (MGSD) for feature representation and a generalized multi-view consistent discriminant metric learning method (GMDA-RC) for metric learning.

and efficient through person Re-ID experiments on three public datasets.

2. Brief review of related work

For the many mathematical notations, a list with brief description is showed as follows.

$\psi_{\mu,\theta}(x,y)$	Gabor filters
$G_{\mu}(x,y)$	Average of $G_{\mu,\theta}(x,y)$
B_i	Biologically inspired feature (BIF) magnitude images
$GC_i(x,y)$	Granularity color images
$G\theta'_i(x,y)$	Granularity orientation images
S_W, S_B, S_C, S_M	Within-class variations, between-class variations, canonical correlation matrix, representation-consistency scatter matrix
S_{jr}, D_{jr}, C_{jr}, M_{jr}	Two-view within-class scatter matrixes, two-view between-class scatter matrix, two-view canonical correlation matrix, two-view representation-consistency scatter matrix
Z_{jr}	Reconstruction coefficient matrix
E_{jr}	Noise matrix
X_j, X_r	The same person with view-j and view-r
W_{jr}	Project matrix from view-j to view-r
$x_{ijk} \in R^d$	The kth person image form the jth view of the ith person of d dimension
c, v	The number of person, the number of views
n_{ij}	The number of person images from the jth view of ith person

2.1. Color difference histogram

The color different histogram (CDH) [43] algorithm has been exploited for image feature representation, analyzing the perceptually uniform color difference between two points under different backgrounds with color and edge orientations. Specifically, CDH focuses on these salient points, which have same values of color and texture orientation after granularity and encoding. Moreover, these points are apart from the center point with d and located on the boundary of rectangle, shown in Fig. 2. For CDH, only these points with same texture orientations and color index values are utilized to calculate the colors difference histogram. Then, the histogram information is obtained to describe the difference of colors with three channels of original image, considering the feature of color, orientation and the spatial association between the pixels and their d-adjacent neighborhoods. As a result, it is robust to the changes of illumination and viewpoint.

For the color image, the center point is P_{C0} and its d-adjacent neighborhoods are $P_{c1}, P_{c2}, \ldots, P_{c8}$. Similarly, for the texture image, the center point is P_{T0} and its d-adjacent neighborhoods are $P_{t1}, P_{t2}, \ldots, P_{t8}$. Based on the theory of CDH, the color image considers the points of P_{c1}, P_{c7} which have the same texture value and the texture image considers the point of P_{t4}, P_{t8} which have the same color value.

2.2. Classic local discriminant analysis

For the discriminant analysis, S_W and S_B are two covariance matrices describing two classes of variations: the within-class variations and the between-class variations [12]. Given a pair of samples x_i and x_j, x_i, $x_j \in R^{P\times P}$, the S_W and S_B could be obtained by:

$$S_W = \frac{1}{|\text{Å}|}\sum_{x_i,x_j\in \text{Å}} (x_i - x_j)(x_i - x_j)^T \tag{1}$$

$$S_B = \frac{1}{|\mathcal{B}|}\sum_{x_i,x_j\in \mathcal{B}} (x_i - x_j)(x_i - x_j)^T \tag{2}$$

where Å is the similar sample pairs and $\mathcal{B}$ is the dissimilar sample pairs.

To maximize the scatter matrix of between-class and reduce the differences of within-class, the traditional local discriminant analysis would be interested in estimating $K \le k$ column vectors, and obtain the project matrix of $W = [w_1, w_2, \ldots, w_k] \in R^{p\times k}$. It defines the objective optimization function as follows:

$$J(W) = \arg\max_W \frac{trace(W^T S_B W)}{trace(W^T S_W W)} \tag{3}$$

The maximization of $J(W)$ is equivalent to

$$J(W) = \arg\max_W trace(W^T S_B W),\ \ s.t. W^T S_W W = I \tag{4}$$

which can be derived by eigen-decomposition of $S_W^{-1}S_B$.

2.3. Canonical correlation analysis

Canonical correlation analysis (CCA) [45] is an approach to correlating linear relationships between two different views. It aims to obtain two projection directions w_x and w_y for each view and maximize the correlation between the two transformed feature sets in a common subspace. In general, the optimal objective function is defined as follows:

$$\begin{aligned} &(w_x, w_y) = \arg\max_{w_x,w_y} w_x^T C_{xy} w_y \\ &\text{s.t. } w_x^T C_{xx} w_x = 1, w_y^T C_{yy} w_y = 1 \end{aligned} \tag{5}$$

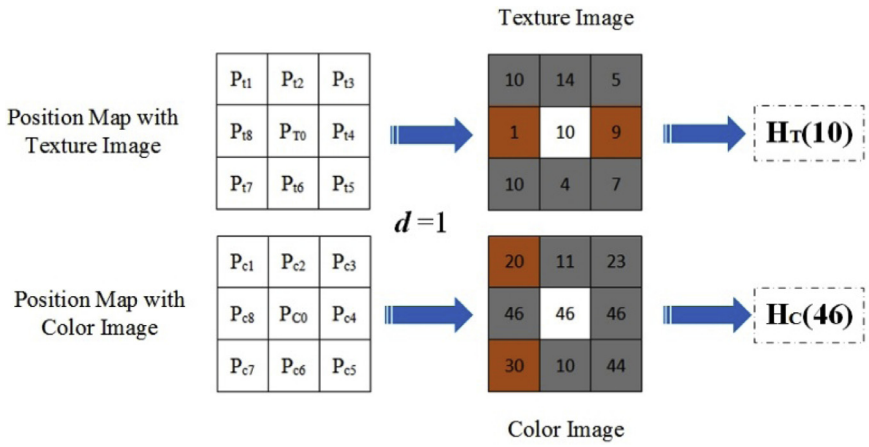


Fig. 2. The processing of Color Difference Histogram (CDH) with $d = 1$.

$$C_{xy} = \frac{1}{n}XY^T, C_{xx} = \frac{1}{n}XX^T, C_{yy} = \frac{1}{n}YY^T \tag{6}$$

where the term $(1/n)$ in Eq. (6) can be canceled out when calculating the correlation coefficient. The Eq. (6) can be transformed into a generalized eigenvalue problem as

$$\begin{bmatrix} \mathbf{0} & C_{xy} \\ C_{yx} & \mathbf{0} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \lambda \begin{bmatrix} C_{xx} & \mathbf{0} \\ \mathbf{0} & C_{yy} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} \tag{7}$$

where $\mathbf{0}$ represents the appropriate number of zero elements.

After surveying related research works, we are motivated to advocate a maximal granularity structure descriptor (GMSD) and a generalized multi-view discriminant analysis with representation consistency (GMDA-RC) for person Re-ID.

3. Maximal granularity structure descriptor

Psychophysical and neurobiological studies have shown that the color and texture, containing rich visual information, play critical roles for human visual perception, and have been widely utilized for image feature representation [43]. However, the illumination conditions and viewpoint changes would lead to vary largely differences of color and texture for the same person. For example, Fig. 16 shows the same person images from Wide Area Re-Identification (WARD) dataset [49] and the same person may vary largely from different camera views. In this section, we propose a robust descriptor (MGSD) to represent appearance of person.

3.1. The idea of the proposed algorithm

To improve the robustness to illumination changes, we apply multiple Gabor filters, which adapt to the habit of observing natural scene with human eyes to preprocess the original images. It is insensitive to illumination changes by analyzing color and texture feature from the multi-scale and multi-orientation. Furthermore, we exploit MAX operator to fuse these images and utilize a sliding window to capture local color difference histogram that considers the salient color and texture features. Moreover, we analyze the horizontal occurrence of local features [12], and maximize the occurrence to make a stable representation against viewpoint and illumination changes. The processes of the proposed method MGSD is shown in Fig. 3.

3.2. Pre-processing with Gabor filters and MAX operator

Gabor filters could reflect the features of the local regions and represent the images by different granularities with multi-scale and multi-orientation [44]. The pre-processing of images with multiple Gabor filters would bring more critical texture information with multi-granularity into our feature representation. To retain more color information, the pre-processing would deal with the three channels (*HSV*) of images respectively, rather than a gray image. Then, we define the multiple Gabor filters as follows:

$$\psi_{\mu,\theta}(x,y) = \frac{\mu^2}{\sigma^2} e^{-\frac{\mu^2\left(x^2+y^2\right)}{2\sigma^2}} \left(e^{i\mu(x\cos\theta + y\sin\theta)} - e^{-\frac{\sigma^2}{2}} \right) \tag{8}$$

where x and y are the coordinate positions in an image, σ is the standard deviation of the Gaussian function, μ represents the scale which is granulated into 16 different scales, and θ defines the orientation which is granulated into 8 orientations.

We compute the original image $I(x, y)$ with multiple Gabor filters and obtain the $G_{\mu,\theta}(x, y)$, as follows:

$$G_{\mu,\theta}(x,y) = I(x,y) \times \psi_{\mu,\theta}(x,y) \tag{9}$$

Then, we extract the averaged feature maps with different orientations and obtain 16-scale feature maps, defined as follows:

$$G_{\mu}(x,y) = \frac{1}{8}\sum_{\nu=1}^{8} G_{\mu,\theta}(x,y) \tag{10}$$

where the $G_{\mu}(x, y)$ is defined as the average of $G_{\mu,\theta}(x, y)$ and the detailed processing is shown in Fig. 4. Furthermore, we divide the 16-scale feature maps $G_{\mu}(x, y)$ into 8 groups and each group includes two neighborhood scale images. Then we take advantage of MAX pooling operator to obtain the biologically inspired feature (BIF) magnitude images in every group, defined as follows,

$$B_i = \max\left(G_{2i-1}(x,y), G_{2i}(x,y)\right), i \in [1,2,\ldots,8] \tag{11}$$

Fig. 5 shows pairs of images and BIF Magnitude Images for an image with three channels of *HSV* color space.

3.3. Maximal granularity structure descriptor

Under different cameras, pedestrians usually appear in different viewpoints that would result in matching same pedestrian difficultly. To address this problem, a novel maximal granularity structure descriptor (MGSD) is proposed, which combines color information, texture orientation and spatial correlation in local regions. In the proposed method, we granulate the *HSV* color space into $4 \times 4 \times 4 = 64$-bins with BIF magnitude images and obtain the granularity color images $GC_i(x, y) \in [1, 2, \ldots, 64]$, $i \in [1, 2, \ldots, 8]$ to reduce the dimensionality of features. Besides, we granulate the color texture image into 36-bins with uniform step and obtain the granularity texture images $G\theta_i'(x, y) \in [1, 2, \ldots, 36]$, $i \in [1, 2, \ldots, 8]$.

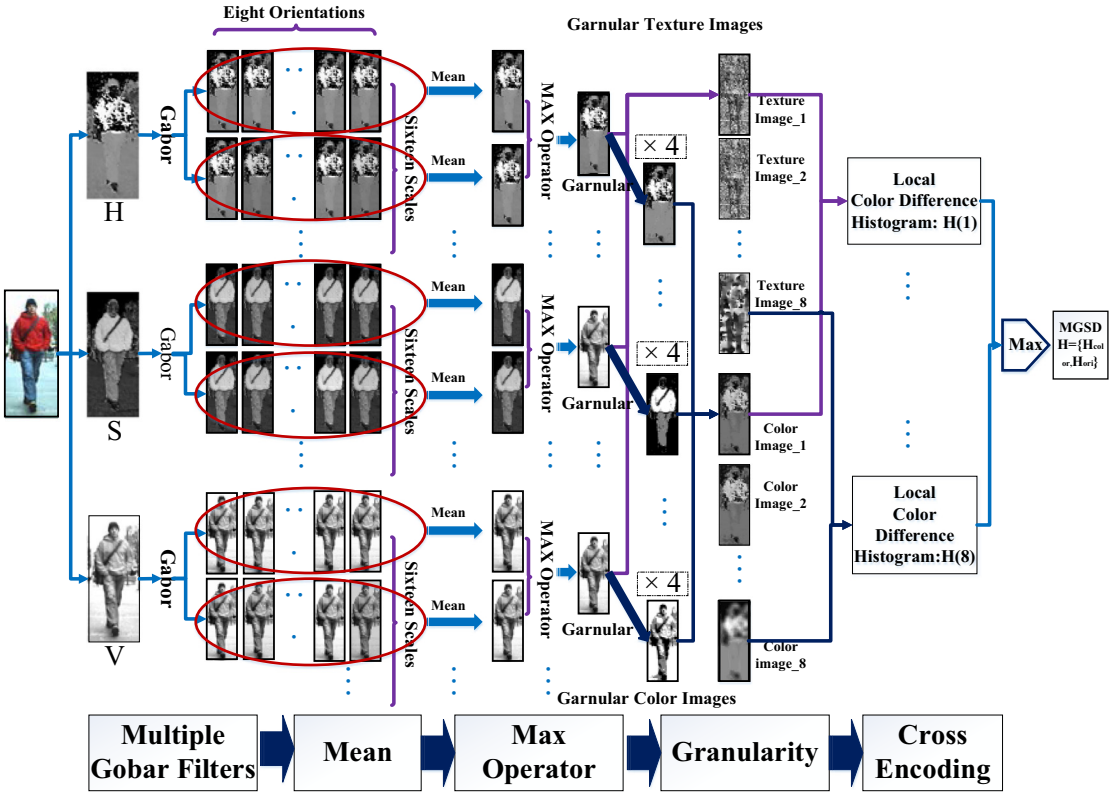


Fig. 3. The integrated framework of the proposed method MGSD.

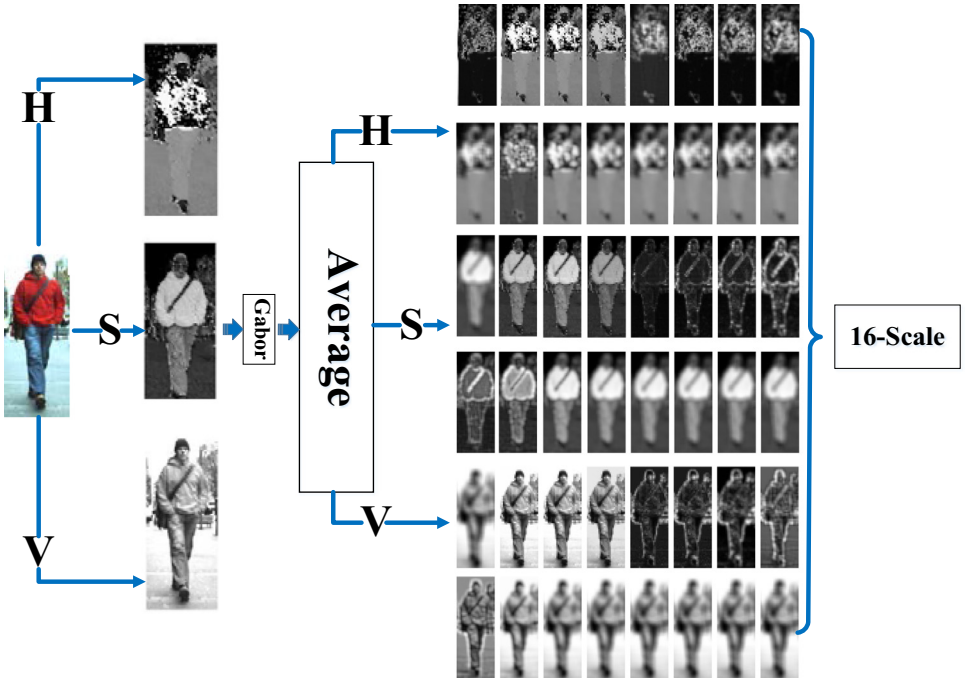


Fig. 4. The framework of pre-processing via multiple Gabor filters with 16 scales and averaged with 8 orientations.

Next, we consider the center pixel and its d-adjacent neighborhoods, and extract the color difference histogram (CDH) [43] that considers the salient points with same color value or texture value, shown in Fig. 6. It is defined as follows:

$$h_{color}^{(i)}(GC_i(\mathrm{x},\mathrm{y})) = \begin{cases} \sum\sum\sqrt{(\Delta H_i)^2+(\Delta S_i)^2+(\Delta V_i)^2} \\ where\ G\theta'_i(x,y) = G\theta'_i(x',y');\ \max\left(|x-x'|,|y-y'|\right) = D_i \end{cases} \tag{12}$$

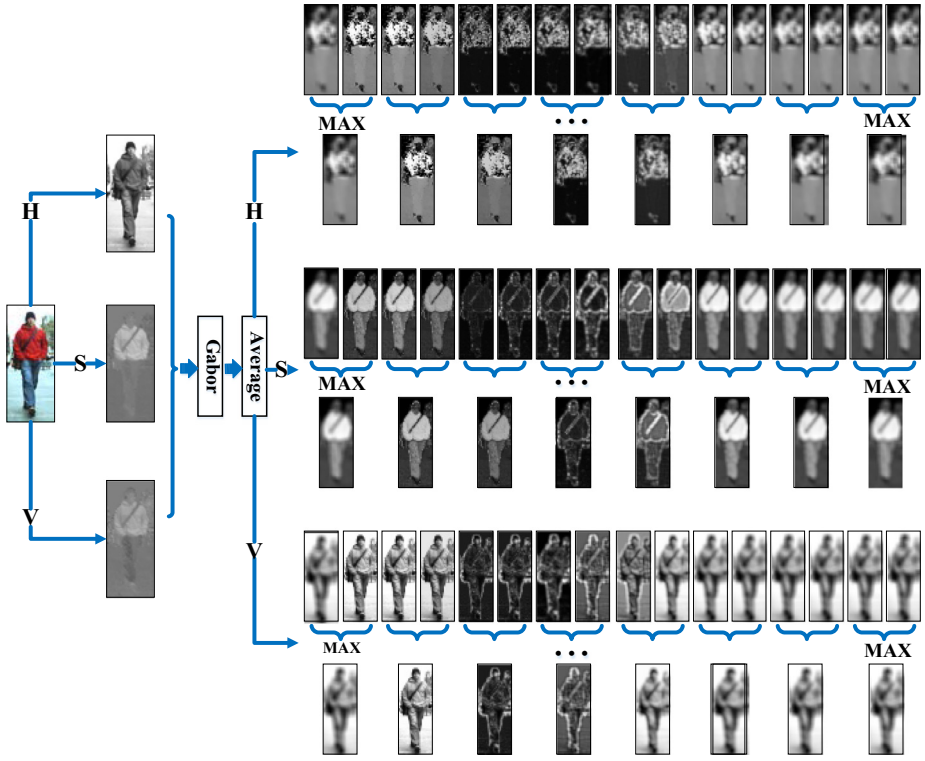


Fig. 5. The pre-processing with MAX pooling operator and we can obtain 8-group feature maps from 16-scale feature maps.

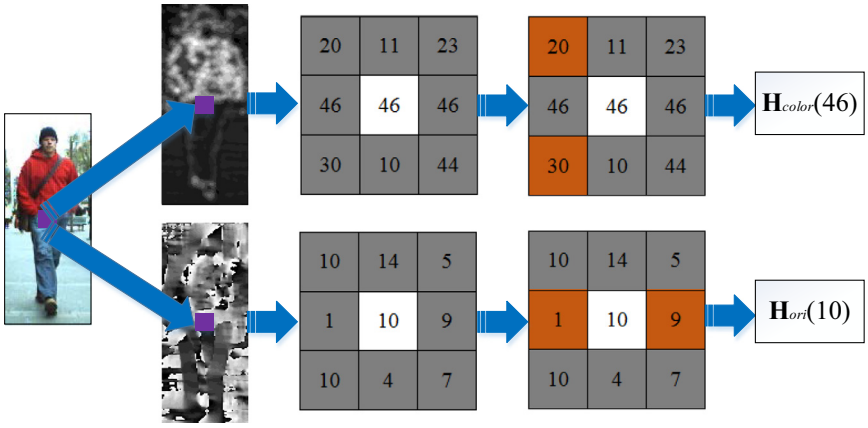


Fig. 6. The processing of local cross encoding: we consider the salient points inspired by color difference histogram, which captures the texture points with same color value or the color points with same texture value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$h_{ori}^{(i)}\left(G\theta_i'(x,y)\right) = \begin{cases} \sum\sum\sqrt{(\Delta H_i)^2+(\Delta S_i)^2+(\Delta V_i)^2} \\ where\ GC_i(x,y) = GC_i\left(x',y'\right);\ \max\left(|x-x'|,|y-y'|\right) = D_i \end{cases} \tag{13}$$

where (x, y) and (x', y') describe the coordinate positions of the points in images. D_i defines the distance between a center point and its neighborhoods.

In order to capture more minutiae features, we exploit a sliding window to capture CDH descriptor in the local regions and consider it as the occurrence probability of one pattern. Then, we choose the maximal values of CDH histograms captured from all local regions at the same horizontal location and obtain the feature vector from the color image C_i(x, y) and texture orientation image $\theta_i'(x, y)$, defined as

$$H = \left[h_{color}^{(i,m)}(0), h_{color}^{(i,m)}(1), \ldots, h_{color}^{(i,m)}(63), h_{ori}^{(i,m)}(0), h_{ori}^{(i,m)}(1), \ldots, h_{ori}^{(i,m)}(35)\right] \tag{14}$$

where m is the number of rows in the images.

Furthermore, we capture the granularity color histogram in *HSV* color space and the scale invariant local ternary pattern (SILTP) [12] from the local regions, and combine them defined as the maximal granularity structure descriptor (MGSD) to represent the person images.

For multi-granularity person images (GC_i(x, y) and $\theta_i'(x, y)$) with Gabor filters and MAX operator, we can capture the final feature vector as follows,

$$MGSD = \left[MGSD_h^{(1)}, MGSD_h^{(2)}, \ldots, MGSD_h^{(i)}, \ldots, MGSD_h^{(8)}\right] \tag{15}$$

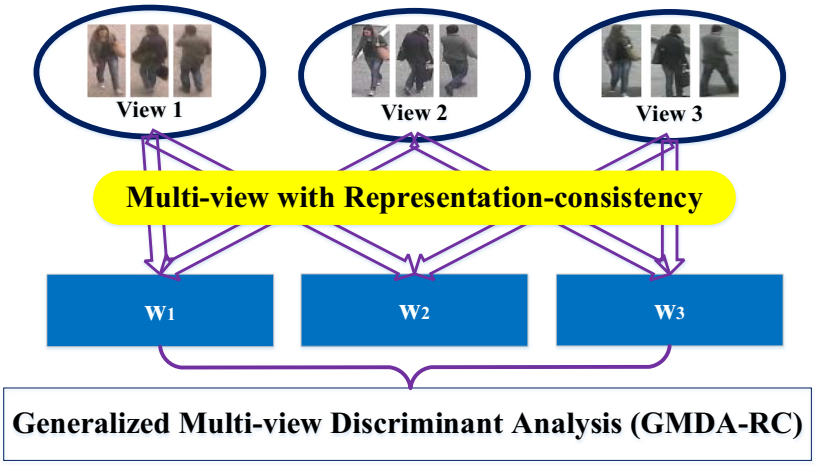


Fig. 7. The process of generalized multi-view discriminant analysis (GMDA-RC) which considers multiple views and capture multiple project matrices.

4. Generalized multi-view discriminant analysis with representation consistency

In this section, we firstly introduce a novel metric learning method based on the formulation of multi-view discriminant analysis and multi-view canonical correlation analysis, and then present its analytic solution, defined as Generalized Multi-view Discriminant Analysis with Representation Consistency (GMDA-RC). It can learn multiple view discriminant subspaces $W = \{w_1^*, w_2^*, \ldots, w_v^*\} \in R^{d\times r}$ with representation consistency and is different from the general framework achieving a discriminative common subspace for all views [45–47,58,59]. Then, we describe its formulation and analytic solution. Note that, the proposed GMDA-RC considers view label information in the multi-view problem of person Re-ID and solve it analytically through generalized eigenvalue decomposition.

4.1. Multi-view discriminant analysis

As shown in Fig. 7, our proposed GMDA-RC attempts to find v linear transforms $w_1^*, w_2^*, \ldots, w_v^*$ that can respectively project the person samples from v different views to an embedding discriminative space, where the between-class variants are maximized and the within-class variants are minimized. In order to consider the view label information, we define $X = \{x_{ijk} | i = 1, 2, \ldots, c;\ j = 1, 2, \ldots, v;\ k = 1, 2, \ldots, n_{ij}\}$ as the set of person images, where $x_{ijk} \in R^d$ is the kth person image form the jth view of the ith person of d dimension, c is the number of person, v is the number of views, n_{ij} is the number of person images from the jth view of ith person. Then, we optimize multiple distinct liner transforms and obtain the two-view within-class scatter matrixes (S_{jr}) [47], considering view label information, defined as follows:

$$S_{jr} = \begin{cases} \sum_{i=1}^{c} \left(\sum_{k=1}^{n_{ij}} x_{ijk} x_{ijk}^T - \frac{n_{ij} n_{ij}}{n_i} \mu_{ij} \mu_{ij}^T \right), & j = r \\ -\sum_{i=1}^{c} \frac{n_{ij} n_{ir}}{n_i} \mu_{ij} \mu_{ir}^T, & otherwise \end{cases} \tag{16}$$

where $\mu_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk}$ is defined as the mean of ith person's images with jth view and n_i is the number of ith person's images.

For the multi-view within-class scatter matrix in the common space, it can be reformulated as follows:

$$S_W = \left[w_1^T, w_2^T, \ldots, w_v^T\right] \begin{pmatrix} S_{11} & \ldots & S_{1v} \\ \vdots & \ddots & \vdots \\ S_{v1} & \ldots & S_{vv} \end{pmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_v \end{bmatrix} = W^T S W \tag{17}$$

Similarly, the two-view between-class scatter matrix can be defined as follows:

$$D_{jr} = \left(\sum_{i=1}^{c} \frac{n_{ij} n_{ir}}{n_i} \mu_{ij} \mu_{ir}^T \right) - \frac{1}{n} \left(\sum_{i=1}^{c} n_{ij} \mu_{ij} \right) \left(\sum_{i=1}^{c} n_{ir} \mu_{ir} \right)^T \tag{18}$$

And we can obtain the multi-view between-class scatter in the common space as follows:

$$S_D = \left[w_1^T, w_2^T, \ldots, w_v^T\right] \begin{pmatrix} D_{11} & \ldots & D_{1v} \\ \vdots & \ddots & \vdots \\ D_{v1} & \ldots & D_{vv} \end{pmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_v \end{bmatrix} = W^T D W \tag{19}$$

Combining with Eq. (3), we can reformulate the multi-view discriminant subspace as follows:

$$(w_1, w_2, \ldots, w_v) = \arg \max_{w_1, \ldots, w_v} \frac{trace\left(W^T D W\right)}{trace\left(W^T S W\right)} \tag{20}$$

which can be solved analytically through general eigenvalue decomposition.

4.2. Multi-view canonical correlation analysis

Different from original multi-view canonical correlation analysis (MCCA) [44], we combine the two-view canonical correlation analysis and define the total correlation in the common space as below:

$$S_C = \left[w_1^T, w_2^T, \ldots, w_v^T\right] \begin{pmatrix} C_{11} & \ldots & C_{1v} \\ \vdots & \ddots & \vdots \\ C_{v1} & \ldots & C_{vv} \end{pmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_v \end{bmatrix} = W^T C W \tag{21}$$

where C_{ij} is defined as Eq. (6), i and j represent two different views. Combining the Eq. (20), we define the generalized multi-

view analysis as follows:

$$(w_1, w_2, \ldots, w_v) = \arg \max_{w_1, w_2, \ldots, w_v} \frac{trace\left(W^T DW\right) + \alpha trace\left(W^T CW\right)}{trace\left(W^T SW\right)} \tag{22}$$

where α is the balance parameter.

4.3. Multi-view representation consistency

For the same person with different camera views, denoted as X_j $(j = 1)$ and $X_r (r = 2)$ respectively, considering the same distributions we can obtain the optimal function based on the consistency of representation as follows:

$$X_j = X_r Z_{jr} + E_{jr} \tag{23}$$

where Z_{jr} is defined as the reconstruction coefficient matrix which can captures the structure of samples. E_{jr} is defined as the noise matrix. However, it cannot ensure the consistency of distributions with different views. Thus, we apply the project matrix (W_{jr}) which is defined the flipping form view-j to view-r, to seek a common subspace to minimize the difference of two views. Then, we can obtain the optimal function as follows:

$$\min_{W_{jr}} \left\| W_{jr}^T M_{jr} W_{jr} \right\|_2^2 \tag{24}$$

$$M_{jr} = \left(X_j - X_r Z_{jr} - E_{jr}\right)\left(X_j - X_r Z_{jr} - E_{jr}\right)^T \tag{25}$$

where j,r represent two-different views, X_j,X_r represent the same person with view-j and view-r. W_{jr} is defined as the project matrix from view-j to view-r. M_{jr} is defined as the two-view representation-consistency scatter matrix for same person with view-j and view-r.

For the Eq. (23), we define $K_{jr} = [Z_{jr}; E_{jr}]$ and $X_r' = [X_r; 1]$. Inspired by the least square method, we can obtain the optimal function of $\frac{1}{2}\|X'_r K_{jr} - X_j\|^2$ and solve it as

$$\begin{cases} X_j = X_r' K_{jr} \\ K_{jr} = \left(X_r'^T X_r'\right)^{-1} X_r'^T X_j \end{cases} \tag{26}$$

The two-view representation-consistency scatter matrix M_{jr} is re-defined as follows,

$$M_{jr} = \left(X_j - X_r' K_{jr}\right)\left(X_j - X_r' K_{jr}\right)^T \tag{27}$$

Then, we can reach the following multi-view representation-consistency scatter matrix:

$$S_M = \left[w_1^T, w_2^T, \ldots, w_v^T\right] \begin{pmatrix} M_{11} & \ldots & M_{1v} \\ \vdots & \ddots & \vdots \\ M_{v1} & \ldots & M_{vv} \end{pmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \ldots \\ w_v \end{bmatrix} = W^T M W \tag{28}$$

Furthermore, combining with Eq. (22), we can define the resemblance as view-consistency for multi-view problem, modeled by the following term:

$$(w_1, w_2, \ldots, w_v) = \arg \max_{w_1, w_2, \ldots, w_v} \frac{trace\left(W^T DW\right) + \alpha trace\left(W^T CW\right)}{trace\left(W^T SW\right) + \beta trace\left(W^T MW\right)} \tag{29}$$

where α, β are the balance parameters.

Through optimizing Eq. (29), the correlation between different views, the discrimination of each view can be maximized simultaneously, and the consistency of representation can been ensured to minimize the differences of distributions with multiple views. We denote this extended multi-view discriminant analysis and multi-view canonical correlation analysis with representation consistency as generalized multi-view discriminant analysis with representation consistency (GMDA-RC).

The objective of GMDA-RC in Eq. (29) can been rewritten as a trace ratio form:

$$(w_1, w_2, \ldots, w_c) = \arg \max_{w_1, w_2, \ldots, w_c} \frac{trace\left(W^T (D + \alpha C) W\right)}{trace\left(W^T (S + \beta M) W\right)} \tag{30}$$

Differing from the conventional discriminant analysis described as Eq. (4), the proposed method i.e. GMDA-RC fuses the multi-view consistency described by matrix M and multi-view canonical correlation described by matrix C. It can lead to a more robust solution for *SVD* and is against the changes of multiple views. Evidently, Eq. (30) can also be solved analytically after relaxing to the ratio trace problem as Eq. (4). That is, the corresponding eigenvector of $(S + \beta M)^{-1}(D + \alpha C)$ with *SVD* is the solution of the objective function. For brevity, more details on the solution can refer to XQDA [12]

4.4. Distance function for GMDA-RC

In this paper, we extern the KISSME approaches to multi-view metric learning and consider the multi-view subspace of $W = \{w_i\} \in R^{d \times r}$, where i represents one view, d represents the dimensionality of original features and r represents the dimensionality of subspace. Then, we define the distance function for GMDA-RC as follows:

$$\begin{aligned} d_{GMDA-RC} = & \left(w_p^T x_p - w_p^T x_q\right)^T w_p (D_{pq} + \alpha C_{pq} - S_{pq} - \beta M_{pq}) \\ & \times w_p^T \left(w_p^T x_p - w_p^T x_q\right) + \left(w_q^T x_p - w_q^T x_q\right)^T w_q \\ & \times (D_{qp} + \alpha C_{qp} - S_{qp} - \beta M_{qp}) w_q^T \left(w_q^T x_p - w_q^T x_q\right) \end{aligned} \tag{28}$$

where x_p, x_q represent the samples with view-p and view-q, w_p, $w_q \in W$, $\boldsymbol{D_{pq}}$ represents the two-view between-class scatter matrix with view-p and view-q, $\boldsymbol{C_{pq}}$ represents the two-view canonical correlation matrix with view-p and view-q, $\boldsymbol{S_{pq}}$ represents the two-view within-class scatter matrix with view-p and view-q, and $\boldsymbol{M_{pq}}$ represents the representation-consistency scatter matrix.

The process of our proposed method (MGSD+GMDA-RC) is shown as follows:

Algorithm 1 The proposed method of MGSD.

Input: dataset $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_m$ with view-1, view-2, ... ,view-m.
Begin:
(1) Pre-processing via multiple Gabor filters with the 16 scales and 8 orientations by Eqs. (8) and (9).
(2) Extract the averaged feature with different orientations and obtain 16-scale feature maps ($\boldsymbol{G_\mu(x, y)}$) by Eq. (10).
(3) Obtain the biologically inspired feature (BIF) magnitude images ($\boldsymbol{B_i}$) by Eq. (11).
(4) Extract the color difference histogram (CDH) by Eq. (12).
(5) Capture the local maximal cross encoding histogram.
(6) Combine the descriptor of local maximal cross encoding histogram, HSV color histogram and SILTP and define it as MGSD.
(7) Obtain final feature vector:
$\boldsymbol{MGSD} = [\boldsymbol{MGSD_h^{(1)}}, \boldsymbol{MGSD_h^{(2)}}, \ldots, \boldsymbol{MGSD_h^{(i)}}, \ldots, \boldsymbol{MGSD_h^{(8)}}]$.
End
Output: MGSD

Algorithm 2 The proposed method of GMDA-RC.

Input: feature matrices $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_m$ with view-1, view-2, ... ,view-m, parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$.
Begin:
(1) Obtain the two-view within-class scatter matrixes ($\boldsymbol{S}_{jr}$) by Eq. (16).
(2) Obtain the multi-view within-class scatter matrix $\boldsymbol{S}_W$ by Eq. (17).
(3) Obtain the two-view between-class scatter matrix $\boldsymbol{D}_{jr}$ by Eq. (18).
(4) Obtain the multi-view between-class scatter $\boldsymbol{S}_D$ by Eq. (19).
(5) Obtain the two-view canonical correlation analysis $\boldsymbol{C}_{jr}$ by Eq. (6).
(6) Obtain the multi-view canonical correlation analysis $\boldsymbol{S}_C$ by Eq. (21).
(7) Obtain the two-view reconstruction coefficient $\boldsymbol{K}_{jr}$ by Eq. (26).
(8) Obtain the two-view representation-consistency scatter matrix $\boldsymbol{M}_{jr}$ by Eq. (25).
(9) Obtain the multi-view representation-consistency scatter matrix $\boldsymbol{S}_M$ by Eq. (28).
(10) Obtain the final optimal function of Eq. (29).
(11) Solve Eq. (29) analytically after relaxing to the ratio trace problem.
End
Output: $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_m$

4.5. Complexity analysis

In our proposed approach, we utilize multiple Gabor filters and capture the local maximal cross encoding histogram. For a region, we can obtain a 100-dimensionality (64-color and 36-texture) histogram. Thus, the dimensionality of proposed descriptor (MGSD) is $100 \times n$, where n is the number of local regions in a person image. For example, we take advantage of overlapping slide window with size of 16×16 to capture local maximal cross encoding histogram 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 $15 \times 100 = 1500$ from a person image with the size of 128×48. Combining the multiple Gabor filters, we can obtain the dimensionality of MGSD is $1500 \times 8 = 12{,}000$. For the metric learning method of GMDA-RC, the main computational complexity is correlative to the process of singular value decomposition (SVD), defined as $t_{svd}(d_{MGSD} \times m)$, where m is the number of views.

5. Experiments

In this section, we show the experiments to evaluate our approach, providing comparisons with the state-of-the-art methods in three publicly available person Re-ID datasets: VIPeR [5], CUHK-01 [12] and Wide Area Re-Identification Dataset (WARD) [49], which cover different aspects and challenges for person Re-ID. In our experiments, we randomly choose all images of p persons (classes) to set up the test set and the rest is test set including a gallery set and a probe set. The gallery set consists of one image for each person and the remaining images are defined as the probe set. This procedure is repeated 10 times.

For evaluation, we utilize Cumulative Matching Characteristic (CMC) curve [50] as the standard performance measurements. The CMC curve represents the expectation of the probe image correct match at rank r against the p gallery images and the rank-1 matching rate is thus the correct matching, recognition rate. In practice, a high rank-1 matching rate is critical, also, a small r value is important because the top matched images will normally be verified by a human operator. In our experiments, we compared the proposed method with several state-of-the-art methods.

In our model, the parameters includes mainly α, β and we obtain the optimal parameters through a method of adjusting one parameter while fixing other parameters [55].

5.1. Experiments on VIPeR dataset

The VIPeR dataset is composed of 1264 images of 632 pedestrians captured by a pair of cameras in an outdoor environment, with two images of 128×48 pixels for every pedestrian. Example images are shown in Fig. 8. The images are taken from horizontal viewpoints but widely different directions. In this dataset, the change of large variations in background, illumination, and viewpoint are the main problems for person Re-ID. In our experiments, we randomly choose 316 pedestrians from all images as training set and the rest is the test set, and repeat the procedure 10 times to get an average performance.

5.1.1. Comparison of the state-of-the-art methods

We compared the proposed method (***MGSD+GMDA-RC***) and (***Fusion of Features+GMDA-RC***) with the state-of-the-art methods including CRAFT [16], GOG [5], LSSL [53], LOMO+MLAPG [54], LOMO+XQDA [12], kLFDA [29], MFA [29], KISSME [51], SVMML [52] and LFDA [28] and reported the results in the Fig. 9 and Table 2. From Fig. 9, we can see that the proposed method of ***MGSD+GMDA-RC*** achieves a recognition rate of 44.87% at rank-1, which can outperform most of the compared methods. Specifically, the ***Fusion of Features*** (i.e. LOMO, GOG, CRAFT) + ***GMDA-RC***

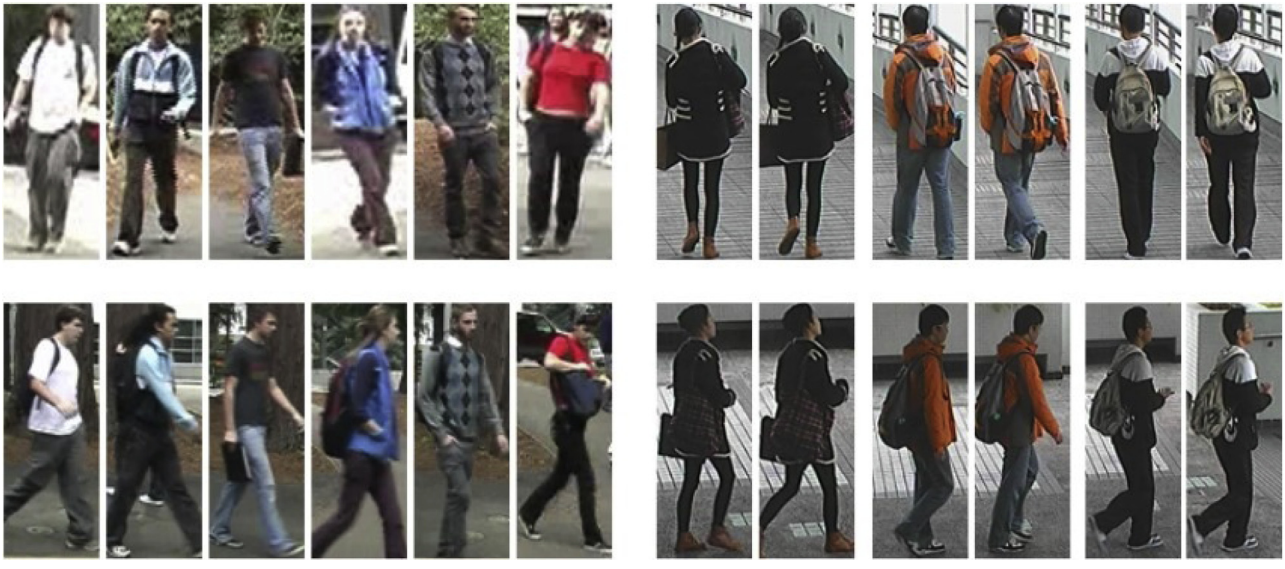


Fig. 8. Examples of the person images in VIPeR dataset for person Re-ID.

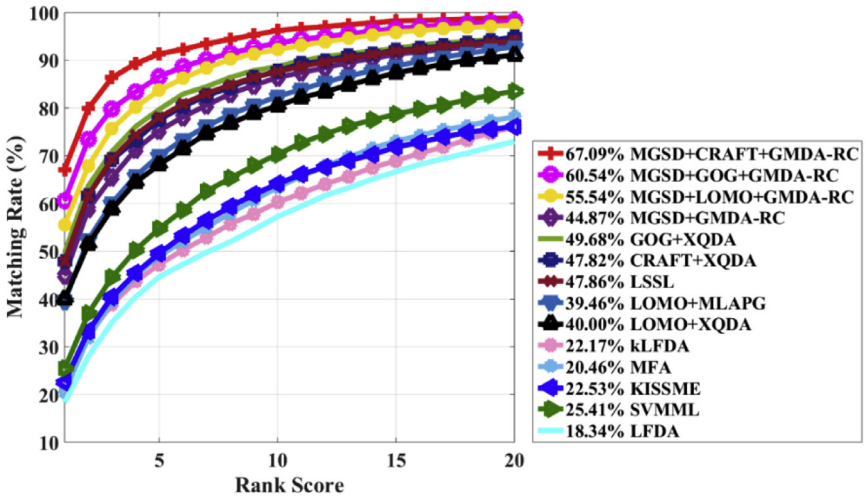


Fig. 9. The CMC curves and rank-1 identification rates on the VIPeR dataset ($P = 316$), by comparing the state-of-the-art methods, including MGSD+GOG+GMDA-RC, MGSD+LOMO+GMDA-RC, MGSD+CRAFT+GMDA-RC, MGSD+GMDA-RC, CRAFT, GOG, LOMO+XQDA, LSSL, rPCCA, kLFDA, MFA, KISSME, SVMML and LFDA.

Table 2
The rank1, 5,10,15,20 matching rate (%) with the methods of MGSD+GOG+GMDA-RC, MGSD+LOMO+GMDA-RC, MGSD+CRAFT+GMDA-RC, MGSD+GMDA-RC, CRAFT, GOG, LSSL, LOMO+MLAPG, LOMO+XQDA, kLFDA, MFA, KISSME, SVMML and LFDA on the VIPeR dataset with $p = 316$. (The bold data is obtained by our methods).

Method	Rank = 1	Rank = 5	Rank = 10	Rank = 15	Rank = 20
MGSD+CRAFT+GMDA-RC	**67.09**	**91.33**	**96.23**	**98.23**	**98.86**
MGSD+GOG+GMDA-RC	**60.54**	**86.58**	**93.61**	**96.68**	**98.20**
MGSD+LOMO+GMDA-RC	**55.54**	**83.70**	**92.25**	**95.89**	**97.37**
MGSD+GMDA-RC	**44.87**	**75.25**	**86.42**	**91.23**	**93.83**
GOG+XQDA	49.68	79.72	88.67	92.50	94.53
CRAFT+XQDA	47.82	77.53	87.78	92.06	94.84
LSSL	47.86	78.03	87.63	91.84	94.05
LOMO+MLAPG	39.46	70.04	82.41	88.83	92.84
LOMO+XQDA	40.00	68.13	80.51	87.37	91.08
kLFDA	22.17	47.23	60.27	68.96	76.01
MFA	20.46	48.97	63.35	73.05	78.15
KISSME	22.53	49.57	64.11	71.79	76.08
SVMML	25.41	54.25	70.28	78.72	83.50
LFDA	18.34	44.64	57.25	66.66	73.96

achieves obvious improvements of (**10.86%, 17.41%, 5.68%**) compared with the best matching rate of (49.68%, GOG) at rank-1. As shown in Table 2, the recognition performance of our approach (***Fusion of Features+GMDA-RC***) is also superior obviously to that of others at rank-5, 10, 20. The intrinsic reasons for significant improvement are as follows. Firstly, we analysis person images from different scales and orientations, and apply MAX operator to extract more salient features which are more conducive to handle illumination variants. Secondly, we capture a local structure and maximize the horizontal occurrence of local features. It is much more stable and effective for the obvious changes of viewpoint. Thirdly, we seek the rich multi-view discriminant information embedded in the person images by maximizing the inter-class variants and minimizing the intra-class variants, considering the multi-view canonical correlation analysis and representation consistency. With these advantages, the proposed method performs better than other methods for person Re-ID.

5.1.2. Comparison of different features

We compared the proposed MGSD with other features (LOMO [9], ELF16 [11], FFN4096 [23], KCCA [17], SCNCD [4]), resulting in the CMC curves and rank-1 matching rates, shown in Fig. 10. For consistency, in the following experiments we utilize the metric learning method of XQDA [12] and GMDA-RC to measure the similarity of the different features. From Fig. 10(a), we can see that the proposed descriptor of MGSD, achieving the match rate of 44.18% with the metric learning method of XQDA, outperforms the other existing descriptors, increasing of 5.18% compared with the method of LOMO+XQDA (39.00%). Furthermore, Fig. 10(b) shows the performance improvement with the proposed metric method of GMDA-RC, achieving the matching rate of 44.87%, and the method of MGSD+GMDA-RC is more significant for person Re-ID. Since these kinds of descriptors are similar in fusing color and texture information, the improvement made by the proposed descriptor of MGSD is mainly due to the specific consideration of minutiae features, which handle the large changes of illumination and viewpoint.

5.1.3. Comparison of metric learning algorithms

We evaluated the proposed GMDA-RC algorithm and several metric learning algorithms, including $l_1 - norm$ distance [27], KISSME [51], kLFDA [29] and XQDA [12], with the same MGSD feature. For the compared algorithms, PCA was first applied to reduce the dimensionality of features. The result of Cumulative Matching Characteristic (CMC) curves is shown in Fig. 11 and we can see that the proposed method (MGSD+GMDA-RC), achieving the match rate of 44.87%, which does better than the other metric learning algorithms, with the accuracy gain reaching 0.69% over the best obtained by (MGSD+XQDA). This indicates that the method of GMDA-

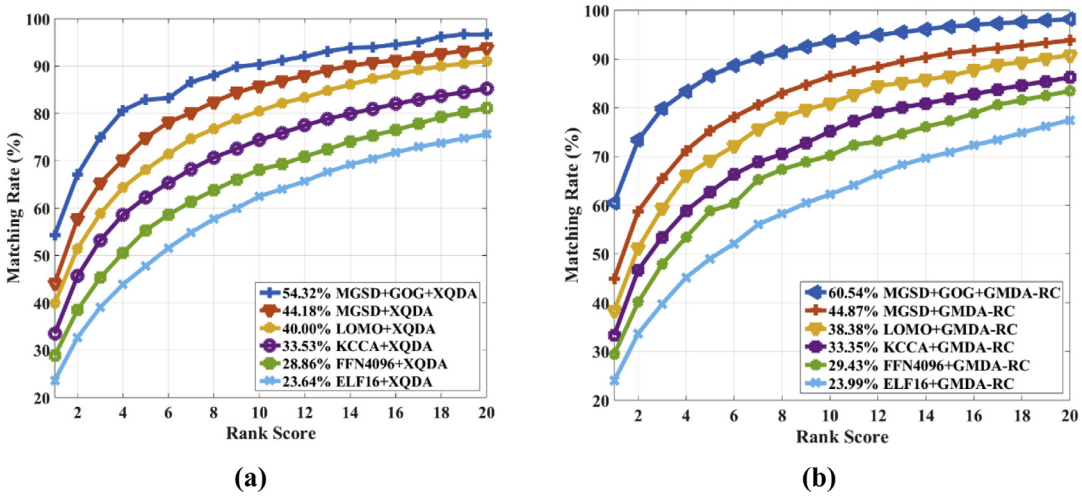


Fig. 10. The CMC curves and rank-1 identification rates on the VIPeR dataset ($P = 316$), by comparing the proposed descriptors of MGSD and MGSD+GOG to four available features including LOMO, KCCA, FFN4096 and ELF16, utilizing the different metric learning methods (XQDA, GMDA-RC).

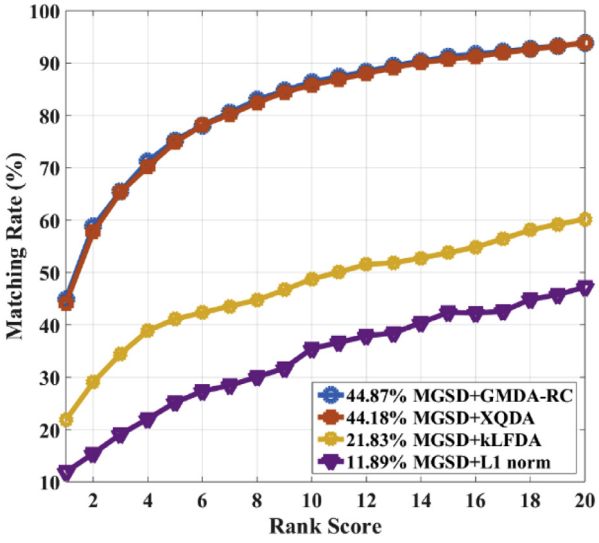


Fig. 11. The CMC curves and rank-1 identification rates on the VIPeR dataset ($P = 316$), by comparing the proposed descriptor of MGSD with GMDA-RC to four available metric learning methods, including $l_1 - norm$ distance, kLFDA and XQDA.

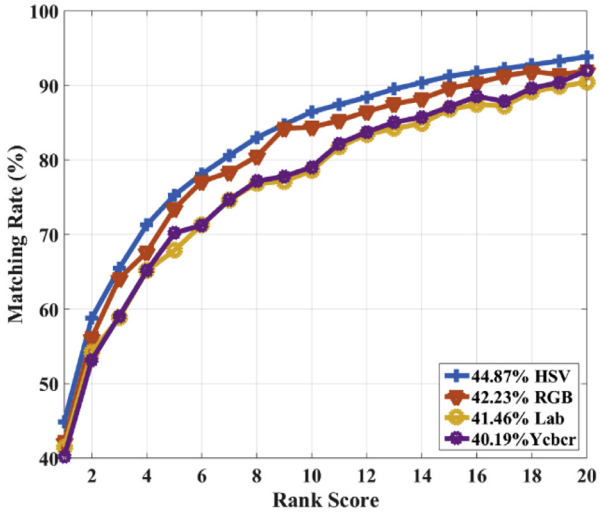


Fig. 12. The CMC curves and rank-1 identification rates of MGSD+GMDA-RC on different color spaces on the VIPeR dataset ($p = 316$).

RC successfully learns multi-view discriminant subspaces as well as an effective metric.

5.1.4. Comparison of different color spaces

Meanwhile, we compare our approach MGSD+GMDA-RC on different color spaces (*HSV, RGB, L*a*b*, Ycbcr*) and the result is shown in Fig. 12. Obviously, in terms of person Re-ID, the MGSD+GMDA-RC on *HSV* color space achieving the rank-1 matching rate of 44.87%, is superior to others with increasing of 2.64%. It is owing to that it contains multiple various channels that represent the images with multiple aspects (e.g. Hue, Saturation, Value) form different granularities.

5.1.5. Comparison of different parameters selection

In this experiment, we compare the performances with different parameters and describe the method of parameters selection. In our model, the parameters include mainly α and β in the proposed method of GMDA-RC. We provide the results of our model with different parameters at rank-1 in Fig. 13. As we can see in this figure, these parameters are not sensitive, performing the best with a small change for person Re-ID. In our model, we obtain the optimal parameters through a method of adjusting one parameter while fixing other parameters, and set the values of α and β as 0.04 and 0.3 respectively. Note that, when $\alpha = 0$ and $\beta = 0.3$, the proposed method of GMDA-RC only considers the influence of Multi-view Representation Consistency (M), reaching the matching rate of 40.96% at rank-1. When $\alpha = 0.04$ and $\beta = 0$, the proposed method of GMDA-RC only considers the influence of Multi-view Canonical correlation analysis (C), reaching the matching rate of 40.32% at rank-1.

5.2. Experiments on CUHK-01 dataset

The CUHK-01 dataset contains 971 persons and 3884 images. Each person has two corresponding images from two different camera views, adding up to four in total. Camera A captures two pedestrian images from the frontal or back view, while camera B captures two from the side view. Meanwhile, the images in this dataset are of high resolution. All images are normalized to 60×160 pixels. In this dataset, the main problem for person Re-ID is the large scale changes in camera view, as shown in Fig. 14. Note

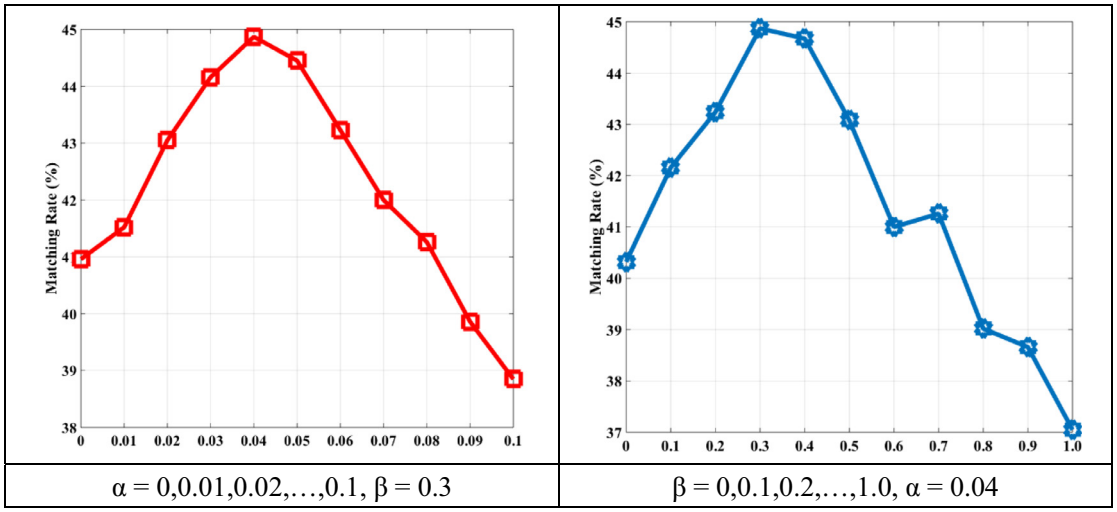


Fig. 13. The performance comparison with different parameters at rank-1.

Fig. 14. Examples of person Re-ID in CUHK-01 dataset.

that we randomly choose 485 persons from all images for training and 486 persons for testing, and randomly select only one image from each person for the gallery set and the rest is the probe set.

We compared our proposed method (MGSD+GMDA-RC) with the state-of-the-art methods, including GOG+XQDA [5], LOMO+MLAPG [54], LOMO+XQDA [12], FFN4096+XQDA [23], kLFDA [29], MFA [29], KISSME [51], SVMML [52] and LFDA [28]. The experimental results are shown in Fig. 15 and Table 3. From the experimental results, we can see that the proposed method significantly outperforms other methods. Besides, it is to be observed that the matching rate of the proposed method (MGSD+GMDA-RC) at rank-1 is up to 70.67%, which exceeds current best result of GOG+XQDA by a margin of 5.34%. And the method of MGSD+CRAFT+GMDA-RC can achieve the best performance of 75.69% at rank-1. Moreover, the CMC curve also reports that the proposed MGSD+GMDA-RC has a better performance than other approaches at rank = 5,10,15,20 on this dataset. Therefore, these results prove that MGSD+GMDA-RC is more robust to the viewpoint and illumination variations, owing to the fact that the project subspace by metric learning method helps to reduce intra-class variations and improves the matching accuracy.

5.3. Experiments on Wide Area Re-Identification dataset (WARD)

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 of particular interest owing to huge illumination variations apart from resolution and pose changes. Example images are shown in Fig. 16. All images are normalized to 48×128 pixels. The dataset has three different cameras and we design the experiments for all the different camera pairs, including 1–2, 1–3 and 2–3. The 70 pedestrians in this dataset are

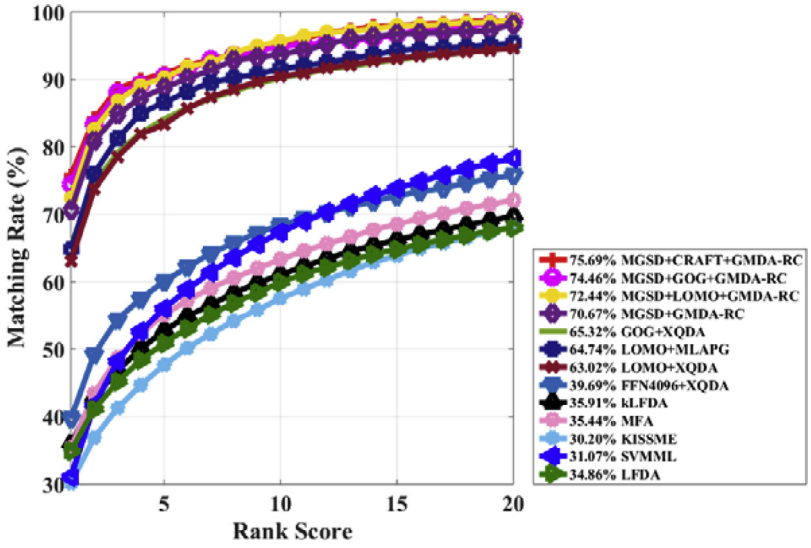
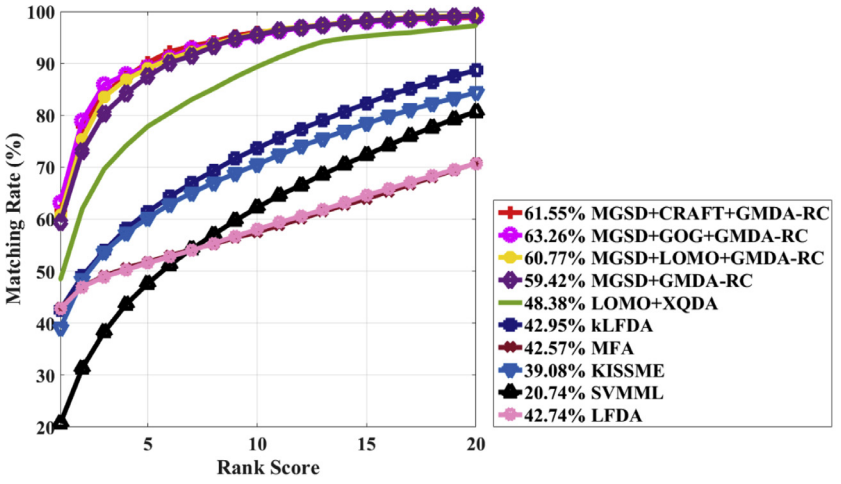


Fig. 15. The CMC curves and rank-1 identification rates on the on the CUHK-01 dataset ($p = 485$), by comparing the state-of-the-art methods, including MGSD+CRAFT+GMDA-RC, MGSD+GOG+GMDA-RC, MGSD+LOMO+GMDA-RC, MGSD+GMDA-RC, GOG+XQDA, LOMO+MLAPG, LOMO+XQDA, FFN4096+XQDA, kLFDA, MFA, KISSME, SVMML and LFDA.

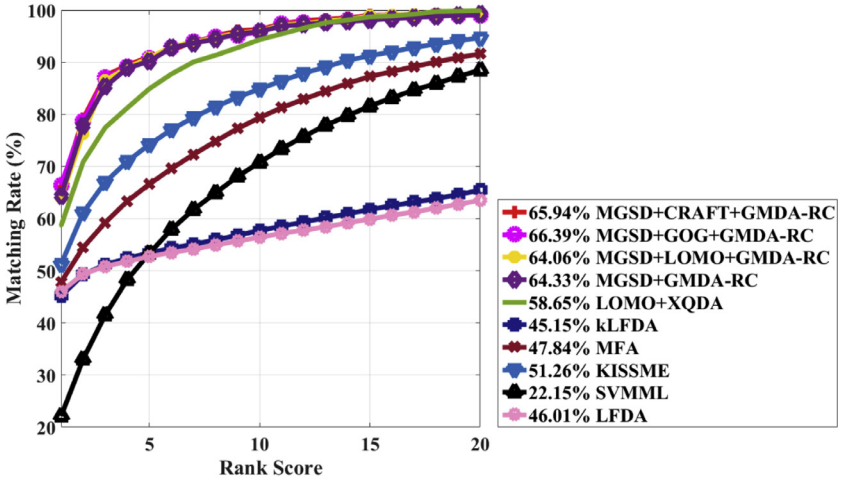
Fig. 16. Examples of person Re-ID in the WARD dataset: (A) the examples acquired by camera A; (B) the examples acquired by camera B; (C) the examples acquired by camera C.

Table 3
The rank 1, 5,10,15,20 matching rate (%) with MGSD+CRAFT+GMDA-RC, MGSD+GOG+GMDA-RC, MGSD+LOMO+GMDA-RC, MGSD+GMDA-RC, GOG+XQDA, LOMO+MLAPG, LOMO+XQDA, FFN4096+XQDA, kLFDA, MFA, KISSME, SVMML and LFDA on the CUHK-01 dataset ($p = 485$). (The bold data is obtained by our methods).

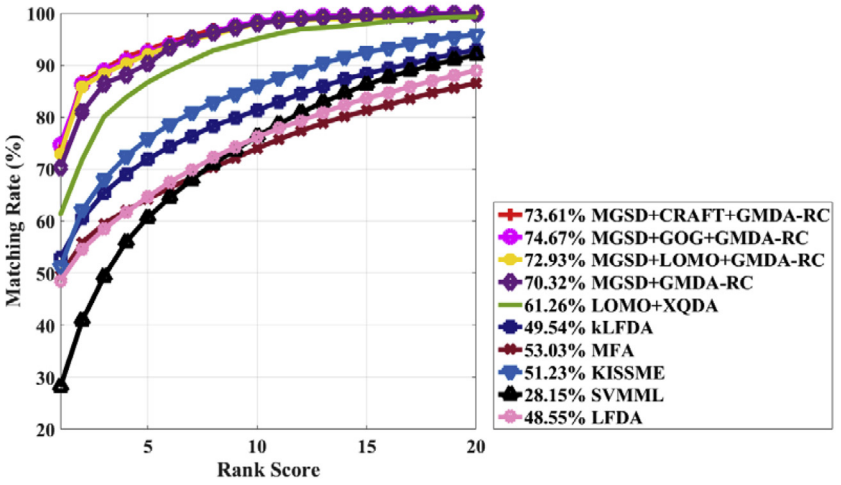
Method	Rank = 1	Rank = 5	Rank = 10	Rank = 15	Rank = 20
MGSD+CRAFT+GMDA-RC	**75.69**	**90.89**	**95.38**	**98**	**98.65**
MGSD+GOG+GMDA-RC	**74.46**	**90.47**	**95**	**96.67**	**98.59**
MGSD+LOMO+GMDA-RC	**72.44**	**89.87**	**95.65**	**98**	**98.8**
MGSD+GMDA-RC	**70.67**	**88.77**	**93.71**	**96.80**	**98.24**
GOG+XQDA	65.33	84.13	90.25	92.94	94.61
LOMO+MLAPG	64.74	86.60	91.55	94.23	95.40
LOMO+XQDA	63.02	83.33	88.87	91.93	93.70
FFN4096+XQDA	39.69	60.05	68.43	72.76	75.79
kLFDA	35.91	52.71	61.05	66.22	69.77
MFA	35.44	55.10	63.3	68.53	72.09
KISSME	30.20	47.66	57.54	63.88	68.16
SVMML	31.07	56.04	67.27	73.83	78.30
LFDA	34.86	50.91	59.91	64.80	68.03



Camrea 1-2



Camera 1-3



(c)camera 2-3

Fig. 17. The CMC curves and rank-1 identification rates on the WARD dataset ($P = 60$) with different camera pairs (e.g. 1–2, 1–3, 2–3), by comparing the state-of-the-art methods, including MGSD+CRAFT+GMDA-RC, MGSD+GOG+GMDA-RC, MGSD+LOMO+GMDA-RC, MGSD+GMDA-RC, LOMO+XQDA, kLFDA, MFA, KISSME, SVMML and LFDA.

Table 4
The rank 1, 5,10 matching rate (%) with the MGSD+CRAFT+GMDA-RC, MGSD+GOG+GMDA-RC, MGSD+LOMO+GMDA-RC, MGSD+GMDA-RC, LOMO+XQDA, kLFDA, MFA, KISSME, SVMML and LFDA on the WARD dataset ($p = 60$). (The bold data is obtained by our methods).

Method	Camera 1–2			Camera 1–3			Camera 2–3		
	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$	$r=1$	$r=5$	$r=10$
MGSD+CRAFT+GMDA-RC	**65.94**	**91.08**	**96.25**	**61.55**	**90.16**	**96**	**73.61**	**93**	**98.52**
MGSD+GOG+GMDA-RC	**66.39**	**90.74**	**95.89**	**63.26**	**89.32**	**95.32**	**74.67**	**92.43**	**98.36**
MGSD+LOMO+GMDA-RC	**64.06**	**90.89**	**96**	**60.77**	**89.01**	**95.82**	**72.93**	**92.08**	**98**
MGSD+GMDA-RC	**64.33**	**90.21**	**95.91**	**59.42**	**87.50**	**95.49**	**70.32**	**90.53**	**98.04**
LOMO+XQDA	58.65	84.93	94.35	48.38	77.89	89.36	61.26	86.70	95.10
kLFDA	45.15	53.44	57.72	42.95	51.78	57.52	49.54	64.27	74.01
MFA	47.84	66.63	79.38	42.57	61.42	73.69	53.03	71.94	81.39
KISSME	51.26	74.23	84.90	39.08	60.24	70.54	51.23	75.85	87.52
SVMML	22.15	53.34	70.08	20.74	57.63	62.24	28.15	60.73	78.68
LFDA	46.01	52.74	56.41	42.74	51.58	58.04	48.55	64.68	77.87

divided into training set which contains 10 persons and testing set which is composed of 60 persons.

We compared our proposed MGSD+CRAFT+GMDA-RC, MGSD+GOG+GMDA-RC, MGSD+LOMO+GMDA-RC, MGSD+GMDA-RC with the state-of-the-art methods including LOMO+XQDA [12], kLFDA [29], MFA [29], KISSME [51], SVMML [52] and LFDA [28]. The experimental results are reported in Fig. 17 and Table 4. For the camera pairs 1–2, 1–3 and 2–3, our method achieves always better results than others. From Table 4, with varying camera pairs (e.g. 1–2, 1–3, 2–3), we can see that the performances of our approach MGSD+GMDA-RC (64.33%, 59.42%, 70.32%) are respectively improved by 5.68%, 11.04%, 9.06% at rank = 1 for camera pairs 1–2, 1–3 and 2–3, compared to the current best result of LOMO+XQDA (58.65%, 48.38%, 61.26%). It indicates that our approach considers the representation consistency of multiple camera views and can effectively adapt to the huge illumination variations apart from resolution and pose changes with multiple camera views. Meanwhile, the CMC curve also proves that out proposed method MGSD+GMDA-RC has a better performance at rank > 1 than other methods. That is to say, our proposed approach has a significantly better performance than other state-of-the-art methods on this dataset. Besides, the combined features (MGSD+LOMO, MGSD+GOG, MGSD+CRAFT) can perform better than MGSD because of more captured information such as multiple color space (GOG) and the camera correlation aware feature augmentation (CRAFT) on this dataset.

6. Conclusion

In this paper, we propose an efficient approach including maximal granularity structure descriptor (MGSD) and metric learning (GMDA-RC) for person Re-ID, which can extract salient information and capture rich discriminative appearance whilst being robust against complex environments. The MGSD is an efficient descriptor and robust against the changes of viewpoint and illumination. In MGSD, we present a novel local maximal horizontal occurrence strategy on the biologically inspired feature (BIF) magnitude images, described by local maximal crossing coding histogram. It can perfectly integrate the color, texture and spatial structural information. Meanwhile, we have also proposed a metric learning approach based on multiple cross-views of discriminant analysis, called GMDA-RC that is very effective in dealing with the curse of dimension problem faced by person Re-ID. Experimental results on three challenging person Re-ID datasets, VIPeR, CUHK-01 and WARD, show that the proposed method i.e. MGSD+GMDA-RC significantly outperforms the state-of-the-art methods. Furthermore, it can be also adapted to other image matching problems, such as the heterogeneous face recognition. The future work will try to address optimal choice of the parameters. In addition, the transfer learning method could solve the problem of inconsistent distributions across multiple views and we will try to extend our model via its theory.

Acknowledgments

The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions. This work is supported by China National Natural Science Foundation under grant No. 61673299, 61203247, 61673301, 61273304. This work is also partially supported by China National Natural Science Foundation under grant No. 61573259, 61573255. It is also supported by the Fundamental Research Funds for the Central Universities (Grant No. 0800219327). 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 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.

References

[1] S.T. Possony, On The Russian Demographic Changes, Defense & Foreign Affairs, 1983.
[2] L. Zheng, Y. Yang, A.G. Hauptmann. Person re-identification: past, present and future. arXiv preprint arXiv:1610.02984, 2016.
[3] K. Igor, A. Amit, R. Ehud, Color invariants for person reidentification, IEEE Trans. Pattern Anal. Mach. Intell. 35 (7) (2013) 1622–1634 12.
[4] Y. Yang, J. Yang, J. Yan, et al., Salient color names for person re-identification, in: European Conference on Computer Vision (ECCV), 2014, pp. 536–551.
[5] T. Matsukawa, T. Okabe, E. Suzuki, et al., Hierarchical Gaussian descriptor for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1363–1372.
[6] C. Zhao, X. Wang, W. Wong, W. Zheng, J. Yang, D. Miao, Multiple metric learning based on bar-shape descriptor for person re-identification, Pattern Recognit. 71 (2017) 218–234.
[7] J. García, A. Gardel, I. Bravo, et al., Multiple view oriented matching algorithm for people reidentification, IEEE Trans. Ind. Inf. 10 (3) (2014) 1841–1851.
[8] Z. Zhang, V. Saligrama, Prism: person reidentification via structured matching, IEEE Trans. Circuits Syst. Video Technol. 27 (3) (2017) 499–512.
[9] W. Li, Y. Wu, J. Li., Re-identification by neighborhood structure metric learning, Pattern Recognit. 61 (2017) 327–338.
[10] W. Lin, Y. Shen, J. Yan, et al., Learning correspondence structures for person re-identification, IEEE Trans. Image Process. 26 (5) (2017) 2438–2453.
[11] D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in: European Conference on Computer Vision (ECCV), 2008, pp. 262–275.
[12] S. Liao, Y. Hu, X. Zhu, et al., Person re-identification by local maximal occurrence representation and metric learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 2197–2206.
[13] N. Martinel, C. Micheloni, G.L. Foresti, Saliency weighted features for person re-identification, in: European Conference on Computer Vision (ECCV), 2014, pp. 191–208.
[14] 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 (CVPR), 2013, pp. 3586–3593.

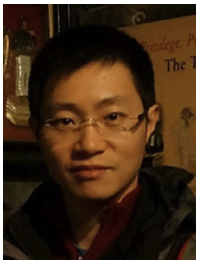
[15] Y.G. Lee, S.C. Chen, J.N. Hwang, et al., An ensemble of invariant features for person reidentification, in: IEEE Transactions on Circuits and Systems for Video Technology, 27, 2017, pp. 470–483.
[16] Y.C. Chen, X. Zhu, W.S. Zheng, et al., Person re-identification by camera correlation aware feature augmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.
[17] G. Lisanti, I. Masi, A. Del Bimbo, Matching people across camera views using kernel canonical correlation analysis, in: Proceedings of the International Conference on Distributed Smart Cameras, 2014, p. 10.
[18] Z. Shi, T.M. Hospedales, T. Xiang, Transferring a semantic representation for person re-identification and search, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4184–4193.
[19] A. Li, L. Liu, K. Wang, et al., Clothing attributes assisted person reidentification, IEEE Trans. Circuits Syst. Video Technol. 25 (5) (2015) 869–878.
[20] C. Su, S. Zhang, J. Xing, et al., Deep attributes driven multi-camera person re-identification, in: European Conference on Computer Vision (ECCV), 2016, pp. 475–491.
[21] Y. Yang, L. Wen, S. Lyu, et al., Unsupervised learning of multi-level descriptors for person re-identification, in: AAAI, 2017, pp. 4306–4312.
[22] T. Xiao, H. Li, W. Ouyang, et al., Learning deep feature representations with domain guided dropout for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1249–1258.
[23] S. Wu, Y.C. Chen, X. Li, et al., An enhanced deep feature representation for person re-identification, in: the IEEE Winter Conference on Applications of Computer Vision (WACV), 2016, pp. 1–8.
[24] A. Franco, L. Oliveira, Convolutional covariance features: conception, integration and performance in person re-identification, Pattern Recognit. 61 (2017) 593–609.
[25] F. Zhu, X. Kong, L. Zheng, et al., Part-based deep hashing for large-scale person re-identification, IEEE Transactions on Image Processing, 2017.
[26] J. Wang, Z. Wang, C. Gao, et al., DeepList: learning deep features with adaptive listwise constraint for person reidentification, IEEE Trans. Circuits Syst. Video Technol. 27 (3) (2017) 513–524.
[27] W.S. Zheng, S. Gong, T. Xiang, Re-identification by relative distance comparison, Pattern Anal. Mach. 35 (3) (2013) 653–668.
[28] S. Pedagadi, J. Orwell, S. Velastin, et al., Local fisher discriminant analysis for pedestrian re-identification, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE, 2013, pp. 3318–3325.
[29] F. Xiong, M. Gou, O. Camps, et al., Person re-identification using kernel-based metric learning methods, in: European Conference on Computer Vision (ECCV), 2014, pp. 1–16.
[30] L. Zhang, X. Tao, S.G. Gong, Learning a discriminative null space for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1239–1248.
[31] R. Zhao, W. Oyang, X. Wang, Person re-identification by saliency learning, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2) (2017) 356–370.
[32] D. Tao, Y. Guo, M. Song, et al., 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] P. Peng, T. Xiang, Y. Wang, et al., Unsupervised cross-dataset transfer learning for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1306–1315.
[35] R. Zhang, L. Lin, R. Zhang, et al., Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification, IEEE Trans. Image Process. 24 (12) (2015) 4766–4779.
[36] 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 (CVPR), 2015, pp. 3908–3916.
[37] S. Ding, L. Lin, G. Wang, et al., Deep feature learning with relative distance comparison for person re-identification, Pattern Recognit. 48 (10) (2015) 2993–3003.
[38] H. Shi H, Y. Yang, X. Zhu, et al., Embedding deep metric for person re-identification: a study against large variations, in: European Conference on Computer Vision (ECCV), 2016, pp. 732–748.
[39] C Sun, D. Wang, H. Lu, Person re-identification via distance metric learning with latent variables, IEEE Trans. Image Process. 26 (1) (2017) 23–34.
[40] S. Paisitkriangkrai, C. Shen, A. van den Hengel, Learning to rank in person re-identification with metric ensembles, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1846–1855.
[41] S.Z. Chen, C.C. Guo, J.H. Lai, Deep ranking for person re-identification via joint representation learning, IEEE Trans. Image Process. 25 (5) (2016) 2353–2367.
[42] C.L. Chen, C.Y. Zhang, Data-intensive applications, challenges, techniques and technologies: a survey on Big Data, Information Sciences 275 (2014) 314–347.
[43] G.H. Liu, J.Y. Yang, Content-based image retrieval using color difference histogram, Pattern Recognit. 46 (2013) 188–198.
[44] B. Ma, Y. Su, F. Jurie, Bicov: a novel image representation for person reidentification and face verification, in: Proceedings of the British Machine Vision Conference, 2012, pp. 57.1–57.11.
[45] J. Rupnik, J. Shawe-Taylor, Multi-view canonical correlation analysis, in: Conference on Data Mining and Data Warehouses (SiKDD), 2010, pp. 1–4.
[46] A. Sharma, A. Kumar, H. Daume, et al., Generalized multiview analysis: a discriminative latent space, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2160–2167.
[47] M. Kan, S. Shan, H. Zhang, et al., Multi-view discriminant analysis, IEEE Trans. Pattern Anal. Mach. Intell. 38 (1) (2016) 188–194.
[48] X. Han, L. Clemmensen, Regularized generalized eigen-decomposition with applications to sparse supervised feature extraction and sparse discriminant analysis, Pattern Recognit. 49 (2016) 43–54.
[49] A. Das, A. Chakraborty, A.K. Roy-Chowdhury, Consistent re-identification in a camera network, in: European Conference on Computer Vision, 2014, pp. 330–345.
[50] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recognition reacquisition, and tracking, in: Proc. IEEE Int't Workshop Performance Evaluation of Tracking and Surveillance, 2007.
[51] M. Kostinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, Large scale metric learning from equivalence constraints, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2288–2295.
[52] Z. Li, S. Chang, F. Liang, T.S. Huang, L. Cao, J.R. Smith, Learning locally adaptive decision functions for person verification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3610–3617.
[53] Y. Yang, S. Liao, Z. Lei, S.Z. Li, Large scale similarity learning using similar pairs for person verification, Thirtieth AAAI Conference on Artificial Intelligence, 2016.
[54] S. Liao, S.Z. Li, Efficient PSD constrained asymmetric metric learning for person re-identification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 3685–3693.
[55] M. Valipour, Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms, Meteorol. Appl. 23 (1) (2016) 91–100.
[56] D. Tao, X. Li, X. Wu, S.J. Maybank, Geometric mean for subspace selection, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2008) 260–274.
[57] D. Tao, X. Li, X. Wu, S.J. Maybank, General tensor discriminant analysis and gabor features for gait recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (10) (2007) 1700–1715.
[58] W. Liu, D. Tao, J. Cheng, Y. Tang, Multiview Hessian discriminative sparse coding for image annotation, Comput. Vision Image Understanding 118 (1) (2013) 50–60.
[59] W. Liu, H. Liu, D. Tao, K. Lu, Multiview Hessian regularized logistic regression for action recognition, Signal Process. 110 (5) (2014) 101–107.
[60] J. Li, C. Xu, W. Yang, C. Sun, D. Tao, Discriminative multi-view interactive image re-ranking, IEEE Trans. Image Process. 26 (7) (2017) 3113 A Publication of the IEEE Signal Processing Society.

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 is the author of more than 30 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.

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.

Hanli Wang received the B.E. and M.E. degrees in electrical engineering from Zhejiang University, Hangzhou, China, in 2001 and 2004, respectively, and the Ph.D. degree in computer science from City University of Hong Kong, Kowloon, Hong Kong, in 2007. From 2007 to 2008, he was a Research Fellow with the Department of Computer Science, City University of Hong Kong. From 2007 to 2008, he was also a Visiting Scholar at Stanford University, Palo Alto, CA, USA. From 2008 to 2009, he was a Research Engineer with Precoad, Inc., Menlo Park, CA, USA. From 2009 to 2010, he was an Alexander von Humboldt Research Fellow at University of Hagen, Hagen, Germany. Since 2010, he has been a full Professor with the Department of Computer Science and Technology, Tongji University, Shanghai, China. His current research interests include digital video coding, computer vision, and machine learning.



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.

Yong Xu (M'06SM'15) received the B.S. and M.S. degrees from the Air Force Institute of Meteorology, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology, Nanjing, in 2005. Currently, he is a professor in the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China. His current research interests include pattern recognition, biometrics.

David Zhang (F'08) received the Degree in computer science from Peking University, Beijing, China, the M.Sc. degree in computer science and the Ph.D. degree from the Harbin Institute of Technology (HIT), Shenzhen, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 1994. He was a Post-Doctoral Fellow with Tsinghua University, Beijing, and then an Associate Professor with Academia Sinica, Beijing, from 1986 to 1988. He is currently the Head of the Department of Computing, and a Chair Professor with The Hong Kong Polytechnic University, Hong Kong. He also serves as the Visiting Chair Professor with Tsinghua University, and an Adjunct Professor with Peking University, Shanghai Jiao Tong University, Shanghai, China, HIT, and the University of Waterloo. He has authored over ten books and 200 journal papers.

Prof. Zhang is a Croucher Senior Research Fellow, Distinguished Speaker of the IEEE Computer Society, and fellow of the International Association for Pattern Recognition. He is the Founder and Editor-in-Chief of the International Journal of Image and Graphics, a Book Editor of International Series on Biometrics (Springer), an Organizer of the International Conference on Biometrics Authentication, and an Associate Editor of more than ten international journals, including the IEEE TRANSACTIONS and Pattern Recognition.