Hindawi
Complexity
Volume 2018, Article ID 5940181, 12 pages
https://doi.org/10.1155/2018/5940181

Research Article

Ensemble Learning-Based Person Re-identification with Multiple Feature Representations

Yun Yang,[1,2] Xiaofang Liu,[1,2] Qiongwei Ye,[3] and Dapeng Tao[4]

[1] *National Pilot School of Software, Yunnan University, Kunming, China*
[2] *Kunming Key Laboratory of Data Science and Intelligent Computing, Kunming, China*
[3] *School of Business, Yunnan University of Finance and Economics, Kunming, China*
[4] *School of Information and Engineering, Yunnan University, Kunming, China*

Correspondence should be addressed to Qiongwei Ye; yeqiongwei@163.com

Received 24 April 2018; Revised 4 July 2018; Accepted 15 July 2018; Published 4 September 2018

Academic Editor: Jing Na

As an important application in video surveillance, person reidentification enables automatic tracking of a pedestrian through different disjointed camera views. It essentially focuses on extracting or learning feature representations followed by a matching model using a distance metric. In fact, person reidentification is a difficult task because, first, no universal feature representation can perfectly identify the amount of pedestrians in the gallery obtained by a multicamera system. Although different features can be fused into a composite representation, the fusion still does not fully explore the difference, complementarity, and importance between different features. Second, a matching model always has a limited amount of training samples to learn a distance metric for matching probe images against a gallery, which certainly results in an unstable learning process and poor matching result. In this paper, we address the issues of person reidentification by the ensemble theory, which explores the importance of different feature representations, and reconcile several matching models on different feature representations to an optimal one via our proposed weighting scheme. We have carried out the simulation on two well-recognized person reidentification benchmark datasets: VIPeR and ETHZ. The experimental results demonstrate that our approach achieves state-of-the-art performance.

1. Introduction

Person reidentification aims to recognize and associate a target pedestrian at different occasions after having previously appeared in several cameras with nonoverlapping views. Rather than manual searching, such a reidentification system can intelligently identify targets from images or videos taken from a different camera, and thus has been commonly used in surveillance, security, forensics, healthcare, robotics, retail, transportation, and so on. Person reidentification has become increasingly popular in the community due to its application and research significance. Therefore, many researchers have studied this topic from different aspects of feature level and measurement level, and proposed a variety of approaches to improve the performance of human identity systems. However, they still face many challenges in practical applications: (a) chaotic public scenes, similar pedestrian characteristics, and obstructed pedestrians and (b) obvious changes in appearance due to different lighting conditions, camera parameters, body posture, and so on. In order to solve the above problems, researchers were committed to (1) find out the optimal feature representations and (2) develop robust matching models for promising accuracy.

Feature representation is a fundamental and important part of the person reidentification system. Low-level features [1] such as shape, color, and texture are usually easily and reliably used in visual recognition. Such features are normally encoded into fixed-length format, for example, in the form of histograms [2], covariances [3, 4], or fisher vectors [5], which constitute simple but efficient feature representations. Color histograms have been studied in association with various color spaces [6], and such feature representation is normally robust to the changes of photometric settings of cameras and lighting condition. However, color-based feature representations alone do not have enough discriminative power to deal with a large number of pedestrians with a similar appearance. Therefore, it also needs to exploit the target pedestrian images with other more prominent features such as texture and shape. Shape context constitutes a global shape feature representation. It uses log-polar coordinates to obtain the relative distribution of points in the plane relative to each point on a detected edge or contour. On the other hand, the histogram of oriented gradients (HOG) captures the local details of shapes by constructing histograms of the gradient directions on densely and uniformly spaced cells. Normal shape-based feature representations are sensitive to the changes of poses and variations of viewpoints, hence a local photometric normalization is always applied to the histograms for improved accuracy. On the other hand, some texture filters and descriptors are also used in person reidentification [7], such as the Gabor filter [8, 9] and other linear filter banks [2], color SIFT [10, 11], LBP [12], and region covariance [12, 13]. Some works [14–18] have also studied how to obtain low-level features with high discrimination directly from data. Moreover, deep learning approaches [19–22] have been applied to person reidentification. Although a number of feature representations have been proposed from different perspectives, the results of the study show that no single feature representation can perfectly describe miscellaneous pedestrian images under different visual conditions. Therefore, it is more likely that the aforementioned low-level features need to be concatenated into a composite representation with high dimensionality [23], and consequently, many dimension reduction techniques [24, 25] have to be employed to retain the most effective feature representations for subsequent matching. LFDA [26] is one of the best dimensionality reduction algorithms in many metric learning methods, which can automatically find a suitable distance transformation matrix to capture different classes of data. Different from LFDA, marginal Fisher analysis (MFA) [27] is proposed as a new supervised dimensionality reduction algorithm by designing two graphs that represent the intraclass compactness and interclass separability.

The matching model is another important part of person reidentification. Given a suitable representation obtained from pedestrian images, the matching model aims to match the probe images against a gallery of pedestrian images by measuring the similarity between different images (e.g., Euclidean or Bhattacharyya), or using some model-based matching procedure [28, 29]. Matching models are generally categorized into unsupervised [5, 15, 30–32] and supervised methods [2, 28, 33–35]. The main purpose of the unsupervised method is to design and extract more robust visual features. This method has its special advantages in that it can be extended to different camera fields without any training process. However, it ignores the role of guidance information. In contrast, a method is considered as a supervised approach if it uses labeled samples to adjust parameters of the model and finds relationships between data and corresponding categories. Common works include KISS [36], LMNN [34], ITML [35], LDML [37], PCCA [38], RankSVM [2], and so on. In general, supervised approaches have better performance due to the effective use of prior knowledge. However, in practical applications, the available labeled training set is still quite limited and expensive, which significantly affects its learning quality. Besides, RVM [2], multiple instance learning [39, 40], and partial least squares (PLS) have also been applied to person reidentification, with the same idea of improving the performance of the matching model.

A robust feature representation should be discriminative for miscellaneous pedestrian images under lighting and viewpoint [2, 3, 41], while effective machine learning techniques are essential for the matching model of the reidentification system [10]. There have been many algorithms that have made efforts in the above two directions. Actually, few of the studies have been proposed to reconcile different features and combine the multiple matching results into an optimal solution simultaneously. In this paper, we therefore propose an effective solution by exploiting the difference, complementarity, and importance between different features via the proposed weighting scheme, and the matching model using the RS-KISS distance metric is individually learned on each of selected feature representations. Finally, multiple matching results obtained from different features are combined into an optimal one via the weighted ensemble learning approach.

The main innovations and contributions of this paper are as follows: First, we develop a robust reidentification method to overcome the deficiencies of a single feature representation or composite representation of concatenating multiple features by merging multiple feature representations via an ensemble framework. Second, a novel weighting scheme is proposed to optimally reconcile multiple matching results into an optimal consensus solution, where the weights evaluate the importance of each feature, and take full advantage of the complementarity between different features via the training process. Formal analysis of deriving the weights has also been carried out. Finally, for evaluating the effectiveness and efficiency of our algorithm, we conducted experiments on person reidentification benchmark datasets, and experimental results show that our approach achieves state-of-the-art performance.

The rest of this paper is organized as follows: Section 2 describes the details of our proposed model and its modules. Section 3 reports the testing and simulation results on VIPeR and ETHZ benchmarks. Section 4 discusses the reported experimental results and the issues related to our approach. Finally, the conclusion is drawn in Section 5.

2. Our Approach

In this section, we systematically describe the weighted ensemble model on different feature representations for person reidentification, which includes multiple feature representations, RS-KISS distance metric learning [42], and weighted ensemble learning approach.

2.1. Model Description. The method proposed in this paper is mainly composed of three important parts, including feature representation, parameter learning, and weighted ensemble of matching modules. In our approach, the identification procedure on the size of the target dataset

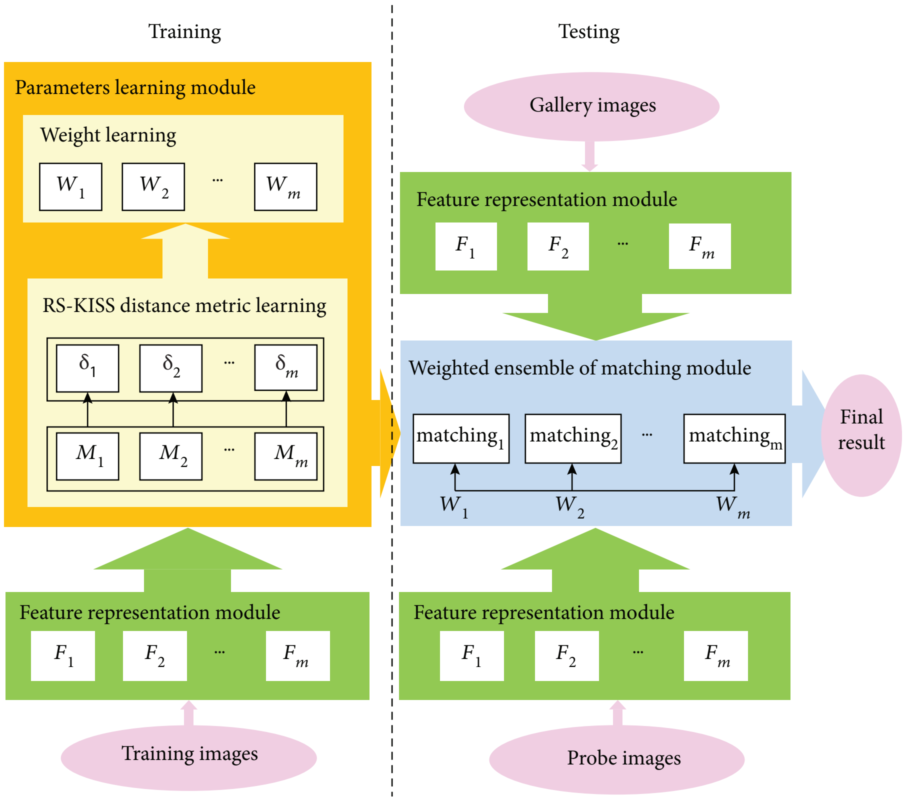


FIGURE 1: Person reidentification model by weighted ensemble learning with multiple feature representations.

as n incurs the computational complexity of $O(n^2)$. As shown in Figure 1, the main functions of each model will be described as follows:

(1) The feature representation module plays an important role as the basis of the entire model. In this part, different features are extracted from the original pedestrian image and constitute a complementary set of feature representations. This will provide the necessary support for the subsequent ensemble learning module, the different features. The features used in this module are described in Section 2.2.

(2) In the parameter learning module, the RS-KISS metric is initially learned by estimating its covariance matrix on the training set as described in Section 2.3, then a novel weighting scheme is proposed to access the importance of different feature representations according to its discriminative power, which is correspondingly defined as a normalized ratio between interperson distance and intraperson distance on the training set. Intuitively larger weights should be assigned to the matching models on the better feature representations.

(3) In the weighted ensemble of the matching module, we integrated the matching results on different features via the ensemble learning schema so that the result fully takes into account the difference and complementarity of different features, and finally combines them into the optimal solution. This is described in detail in Section 2.4.

2.2. Feature Representations. In general, low-level features such as color and texture features are commonly used in person reidentification due to the fact that these features have good generalization capability to different camera views. In our simulation, we thus exploit three color models of red-green-blue (RGB), hue-saturation-value (HSV), luma-blue difference chroma-red difference chroma (YCbCr), and one texture descriptor of local binary pattern (LBP) to obtain four feature representations in the form of a histogram. A histogram represents the distribution of differently binned pixels in an image. The number of bins in a histogram relates to the number of bits in each pixel of an image.

2.2.1. RGB Color Model. The RGB color model [43, 44] is defined by three chromaticities of the red, green, and blue additive primaries, which can be combined to produce any chromaticity. Thus, the RGB histogram is a combination of three histograms based on the red, green, and blue channels of the RGB color space, which indicates the frequency of occurrence of the corresponding color in the image. The RGB histogram represents a discriminative global feature of the image because each image has a unique RGB histogram. It is robust to the changes in rotation, translation, and scaling. But it sometimes does not provide the correct color information due to the problem of luminance effects.

2.2.2. YCbCr Color Model. The YCbCr color model [45, 46] is defined by three components including a luminosity Y, blue-difference chrominance Cb and red-difference chrominance Cr. In fact, the conversions between different color spaces can be made via translation of the representation of a color from one basis to another. YCbCr intends to construct the image into a luminosity component and chrominance components independently; therefore, color and intensity information can be easily separated by using such a color model, which results in a significant discriminative power for recognizing the complex color images with uneven illumination.

2.2.3. HSV Color Model. The HSV color model is commonly employed to describe the color perceived by a human being. In this color model, color information is carried by H (hue or color depth) referring to red, blue, and yellow in the range of 0 to 360° and S (saturation or color purity) taking value from 0 to 1, while the intensity component is represented by V (value or color brightness) in the range of 0 to 1. Although the HSV color model is commonly used for color-based detection and color analysis, the transformation from RGB to HSV is quite time consuming, and if there is a dramatic change in the value of the color information (hue and saturation), pixels with small and large intensities are not considered in such color space.

2.2.4. LBP Descriptor. The LBP descriptor stands for local binary pattern. It is an effective operator for representing the local texture feature that is centered on a pixel as well as local information near the pixel and then results in a binary number. The local binary pattern is a feature description that has received a lot of attention from both the research community and industry due to its two important properties: (1) LBP operator is robust to monotonic gray-scale changes caused by illumination variations and (2) it is possible to analyze images in challenging real-time settings because of its computational simplicity.

In our simulation, the RGB histogram, YCbCr histogram, and HSV histogram are extracted from overlapping blocks with a size of 8×16 and a stride of 8×8 on each image, which encoded the different color distribution information in different color spaces, respectively. On other hand, LBP descriptors are used to extract texture features.

2.3. RS-KISS Distance Metric Learning. RS-KISS is a modified version of the KISS metric [36] by introducing both a smoothing technique [47] and a regularization technique [48], where a robust estimation of covariance matrices can be obtained on a limited amount of training set by averaging the small eigenvalues of a covariance matrix and regulating large eigenvalues of the covariance matrix simultaneously. In such distance metric, with the difference of the feature vector pair $x_{ij} = x_i - x_j$, the distance between x_i and x_j can be measured in (1):

$$\delta\left(x_{ij}\right) = x_{ij}^T\left(\Sigma_1^{-1} - \Sigma_0^{-1}\right)x_{ij}, \tag{1}$$

$$\Sigma_0 = \frac{1}{N_0}\sum_{y_{ij}=0} x_{ij}x_{ij}^T = \frac{1}{N_0}\sum_{y_{ij}=0}\left(x_i - x_j\right)\left(x_i - x_j\right)^T, \tag{2}$$

$$\Sigma_1 = \frac{1}{N_1}\sum_{y_{ij}=1} x_{ij}x_{ij}^T = \frac{1}{N_1}\sum_{y_{ij}=1}\left(x_i - x_j\right)\left(x_i - x_j\right)^T. \tag{3}$$

In (1), the covariance matrices are defined in (2) and (3), where y_{ij} is the label of $x_{ij} : y_{ij} = 1$ as x_i and x_j represent the identical person, otherwise $y_{ij} = 0$. N_0 is the number of similar feature vector pairs, and N_1 is the number of dissimilar feature vector pairs. The above covariance matrix $\boldsymbol{\Sigma}_i$ can be further derived as follows:

$$\boldsymbol{\Sigma}_i = \Phi_i\Lambda_i\Phi_i^T, \tag{4}$$

where λ_{ij} is an eigenvalue of $\boldsymbol{\Sigma}_i$, $\Lambda_i = \text{diag}\,[\lambda_{i1}, \lambda_{i2}, \dots, \lambda_{id}]$, and ϕ_{ij} is an eigenvector of $\boldsymbol{\Sigma}_i$, $\Phi_i = [\phi_{i1}, \phi_{i2}, \dots, \phi_{id}]$.

In RS-KISS, the smoothing technique replaces the first $d - k$ smallest eigenvalues of the estimated covariance matrix $\boldsymbol{\Sigma}_i$ by their average value $\beta_i = 1/d - k\sum_{n=k+1}^{d}\lambda_{\text{in}} \cdot \Lambda_i = \text{diag}\,[\lambda_{i1}, \dots, \lambda_{ik}, \beta_i, \dots, \beta_i]$, while the regularization technique interpolates the covariance matrix $\boldsymbol{\Sigma}_i$ by an identity matrix I. Consequently, it redefines the estimated covariance matrix $\boldsymbol{\Sigma}_i$ shown in (4) as follows:

$$\tilde{\boldsymbol{\Sigma}}_i = (1-\gamma)\boldsymbol{\Sigma}_i + \gamma\alpha_i I = \Phi_i\left\{\text{diag}\left[\left[(1-\gamma)\lambda_{i1} + \gamma\alpha_i, \dots, (1-\gamma)\lambda_{ik} + \gamma\alpha, \underbrace{(1-\gamma)\beta_i + \gamma\alpha_i, \dots, \gamma\alpha, (1-\gamma)\beta_i + \gamma\alpha_i}_{d-k}\right]\right]\right\}\Phi_i^T, \tag{5}$$

where $\alpha_i = (1/d)tr(\boldsymbol{\Sigma}_i)$, and parameter $0 < \gamma < 1$ controls the shrinkage degree of $\tilde{\boldsymbol{\Sigma}}_i$ toward the identity matrix. It practically improves the prediction performance [48, 49] due to the fact that a certain degree of shrinkage resulted in the covariance matrix significantly reduces the effect of its large eigenvalues.

By substituting (5) to (1), we then obtain the RS-KISS distance measure as follows:

$$\begin{aligned}\delta\left(x_{ij}\right) &= x_{ij}^{T}\left(\tilde{\Sigma}_1^{-1}-\tilde{\Sigma}_0^{-1}\right)x_{ij}\\ &= \sum_{n=1}^{k}\frac{1}{(1-\gamma)\lambda_{1n}+\gamma\alpha_1}\left(\phi_{1n}^{T}x_{ij}\right)^2\\ &\quad+\frac{1}{(1-\gamma)\beta_1+\gamma\alpha_1}\left(\left\|x_{ij}\right\|^2-\sum_{n=1}^{k}\left(\phi_{1n}^{T}x_{ij}\right)^2\right)\\ &\quad-\sum_{n=1}^{k}\frac{1}{(1-\gamma)\lambda_{0n}+\gamma\alpha_0}\left(\phi_{0n}^{T}x_{ij}\right)^2\\ &\quad-\frac{1}{(1-\gamma)\beta_0+\gamma\alpha_0}\left(\left\|x_{ij}\right\|^2-\sum_{n=1}^{k}\left(\phi_{0n}^{T}x_{ij}\right)^2\right)\\ &= \left(\frac{1}{(1-\gamma)\lambda_{1n}+\gamma\alpha_1}-\frac{1}{(1-\gamma)\beta_1+\gamma\alpha_1}\right)\sum_{n=1}^{k}\left(\phi_{1n}^{T}x_{ij}\right)^2\\ &\quad+\left(\frac{1}{(1-\gamma)\beta_1+\gamma\alpha_1}-\frac{1}{(1-\gamma)\beta_0+\gamma\alpha_0}\right)\left\|x_{ij}\right\|^2\\ &\quad-\left(\frac{1}{(1-\gamma)\lambda_{0n}+\gamma\alpha_0}-\frac{1}{(1-\gamma)\beta_0+\gamma\alpha_0}\right)\sum_{n=1}^{k}\left(\phi_{0n}^{T}x_{ij}\right)^2.\end{aligned} \tag{6}$$

Eventually, matching or retrieval can be achieved by ranking the individual image x_j from the target gallery based on the above RS-KISS distance $\delta(x_{ij})$ between a probe image x_i and gallery image x_j. As a result, a gallery image with a smaller value of $\delta(x_{ij})$ will be ranked near the top.

2.4. Weighted Ensemble of Matching Models. We have developed several ensemble approaches [50–55] for unsupervised learning tasks. These attempt to improve the robustness of the learning process by combining multiple base learners into a solution, which normally is generally obtained with respect to the average performance of a given individual base learner, leading an effective enabling technique for the joint use of different representations in many pattern recognition systems [56–58]. Although these studies have made significant progress, how to measure the importance of multiple matching results without any a priori information and how to harmonize them together is still a challenging task. To this end, we have introduced a novel weighting scheme that makes the integration process smart and efficient.

2.4.1. Formulation of Weighted Ensemble Approach. Similar to the theoretical framework of the weighted clustering ensemble approach [52], a specific matching result R_i obtained by matching specified probe images or tracks against a target gallery, can be theoretically interpreted as a noisy version of the ground-truth R_c with the same matching task. In other words, the entire solution space can be theoretically constructed from all possible matching results $\mathbf{R}=\{R_i\}$ with normal distribution, and the ground-truth R_c should be the "mean" of all possible matching results:

$$R_C = \arg\min_{R}\sum_{i}\Pr(R_i=R_c)\,d(R_i,R), \tag{7}$$

where Pr $(R_i=R_c)$ is the probability that R_c is randomly distorted to be $\mathbf{R}=\{R_i\}$, and its value is proportional to the similarity between R_i and R_c. $d(\cdot,\cdot)$ is a distance metric.

However, the subset $\{R_m\}_{m=1}^{M}\subseteq\mathbf{R}$ of all possible matching results is normally available in practical situations. The ensemble approach intends to determine an optimal solution by finding the weighted "mean" of M matching results closed to the ground-truth R_c, which was formulized by minimizing the following cost function:

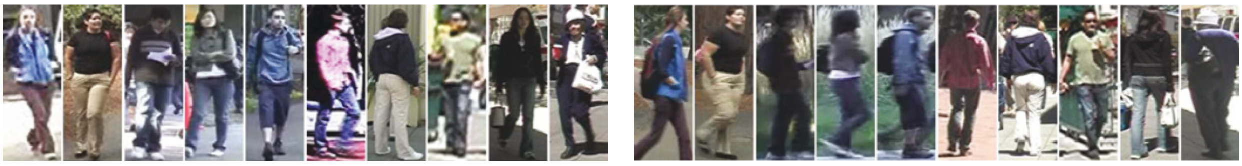
$$J(R)=\sum_{m=1}^{M}w_m d(R_m,R), \tag{8}$$

where $w_m \propto$ Pr $(R_m=R_c)$ and $\sum_{m=1}^{M}w_m=1$. By giving (6), the matching result can be represented as a RS-KISS distance vector $R=[\delta(x_{r1}),\delta(x_{r2}),\dots,\delta(x_{rn})]$, where x_r represents the query target, and x_n represents the reference images in the gallery. Thus (8) can be rewritten as

$$\begin{aligned}J(R) &= \sum_{m=1}^{M}w_m\left\|R_m-R\right\|^2\\ &= \sum_{m=1}^{M}w_m\left\|\left(R_m-R^*\right)+\left(R^*-R\right)\right\|^2\\ &= \sum_{m=1}^{M}w_m\left\|\left(R_m-R^*\right)\right\|^2+2\sum_{m=1}^{M}w_m\left\|\left(R_m-R^*\right)\left(R^*-R\right)\right\|\\ &\quad+\sum_{m=1}^{M}w_m\left\|\left(R^*-R\right)\right\|^2=\sum_{m=1}^{M}w_m\left\|\left(R_m-R^*\right)\right\|^2\\ &\quad+2\left\|\left(R^*-R\right)\left(\sum_{m=1}^{M}w_mR_m-R^*\right)\right\|\\ &\quad+\sum_{m=1}^{M}w_m\left\|\left(R^*-R\right)\right\|^2=\sum_{m=1}^{M}w_m\left\|\left(R_m-R^*\right)\right\|^2\\ &\quad+\sum_{m=1}^{M}w_m\left\|\left(R^*-R\right)\right\|^2.\end{aligned} \tag{9}$$

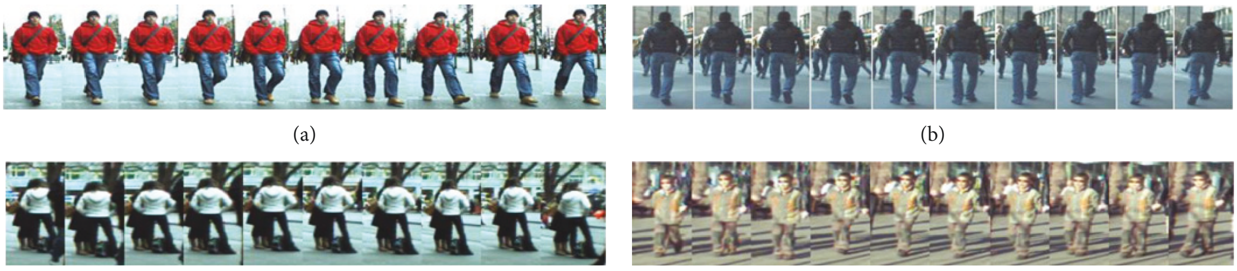
Let $R^*=\sum_{m=1}^{M}w_mR_m$, thus $2\sum_{m=1}^{M}w_m\|(R_m-R^*)(R^*-R)\|=0$ is applied in the evolution of (9), and the actual cost of the weighted ensemble is now decomposed into two terms in the last step.

2.4.2. Weighting Scheme. In (9), the first term corresponds to the quality of multiple matching results, for example, how close they are to the optimal solution, solely determined by the discriminative power of selected feature representations regardless of the ensemble approach. In fact, a feature representation with more discriminative power can determine a better matching result, and theoretically result in a smaller value of $\|(R_m-R^*)\|^2$. Intuitively, the weights could be determined by minimizing the cost of the first term, where larger weights should be assigned to the better matching



(a) (b)

Figure 2: Image examples of the VIPeR dataset.



(a) (b)

(c) (d)

Figure 3: Image examples of the ETHZ dataset.

result obtained on the feature representation of more discriminative power. In our approach, we define the discriminative power of a feature representation as the level of "High cohesion and low coupling", which can be quantified by a normalized ratio between interperson distance and intraperson distance on the training set:

$$w_m = \frac{\left(\sum_{y_{ij}=0}\delta_m\left(x_{ij}\right)+Z\right)/\left(\sum_{y_{ij}=1}\delta_m\left(x_{ij}\right)+Z\right)}{\sum_{m=1}^{M}\left(\left(\sum_{y_{ij}=0}\delta_m\left(x_{ij}\right)+Z\right)/\left(\sum_{y_{ij=1}}\delta_m\left(x_{ij}\right)+Z\right)\right)}. \tag{10}$$

Here, RS-KISS distance is measured with either a positive or negative value, which causes the difficulty in calculating such a ratio. Therefore, we further normalize both interdistance and intradistance into a range of nonnegative values by adding the minimum intradistance $Z = \left|\min\left(\delta_m(x_{ij})\right)\right|_{y_{ij}=1}$ on them, and then the ratio can be calculated in a nonnegative value that appropriately corresponds to the weights of the feature representation.

Once the input matching results $\{R_m\}_{m=1}^{M}$ are obtained for the target dataset, the first term is fixed, and hence the performance of the ensemble approach is primarily controlled by the second term referring to how close the ensemble solution is to the weighted "mean" of the input matching results $\{R_m\}_{m=1}^{M}$. Thus, the optimal solution to minimizing the second term of (9) is obtained as follows:

$$R^* = \arg\min_{R}\sum_{m=1}^{M} w_m d(R_m, R) \Rightarrow \sum_{m=1}^{M} w_m R_m. \tag{11}$$

3. Simulation

In this section, we conducted several experiments to verify the efficiency of our algorithm on the VIPeR [41] dataset and ETHZ [59] dataset, which have been widely used in person reidentification validation. Four feature representations described in Section 2.2 have been used to represent each normalized image in each dataset. The proposed method is also compared with several similar approaches. Given the ground truth, the performance reported in our simulation is significantly better than others.

3.1. Person Reidentification Datasets. The proposed approach has been experimentally validated on two person reidentification datasets (namely the VIPeR and ETHZ datasets). Although it is quite challenging to conduct person reidentification on these datasets since many visual variations including pose changes, viewpoint and lighting variations, and occlusions have to be considered, they has been widely recognized as a benchmark of testing person reidentification approaches. Moreover, it is recognizable to compare our approach with other state-of-the-art techniques on these datasets. In our early work [42], we also conducted the experiment on the iLIDS dataset, but we abstained from using it in this study for fairness. This is because such dataset has many versions available that arbitrarily crops patches from the integrated iLIDS dataset, resulting in various matching results. The image examples of the two selected datasets are shown in Figures 2 and 3, respectively. The details of datasets are given in Sections 3.1.1 and 3.1.2.

3.1.1. VIPeR. VIPeR was established by Gray et al. [41]. It contains 632 pairs of images of persons taken from two different camera views under various conditions. As shown in Figure 2, each intrapersonal image pair is presented in one column, and appearance variation of the same person is mainly caused by a viewpoint change of a certain degrees. Other variations including light conditions, shooting locations, and the image qualities are also considered accordingly.

3.1.2. ETHZ. ETHZ was collected by Ess et al. [59], and was originally proposed for pedestrian detection. Later it was modified for benchmarking person reidentification tasks [60]. The ETHZ dataset has 8580 images collected from 146 subjects. Some example images are shown in Figure 3. Unlike the VIPeR dataset, the ETHZ dataset collects more sample images from each subject as shown in one row of Figure 3. In fact, it consists of three video sequences. The first one has 4857 images of 83 pedestrians, the second one has 1961 images of 35 pedestrians, and third one has 1762 images of 28 pedestrians. It is quite challenging to perform person reidentification tasks on such a dataset due to the illumination changes, scale variations, and occlusions resulting from images of these video sequences.

3.2. Data Preprocessing with Selected Feature Representations. Following the method [36], all of the images are initially normalized to a standard size of 128×64 by dividing the original images into overlapping blocks with a size of 8×16 and a stride of 8×8. Then, HSV, RGB, Ycbcr histograms, and LBP descriptors are extracted from the resized images. As color feature representations, HSV, RGB, and Ycbcr histograms with 24 bins per channel represent the different color distribution information in the HSV, RGB, and Ycbcr color spaces, respectively. As a texture feature representation, the LBP descriptor is used to represent the local information of target images in a binary format. Finally, principal component analysis (PCA) is applied to further reduce the dimensionality of the extracted features in order to accelerate the learning process and remove signal noise.

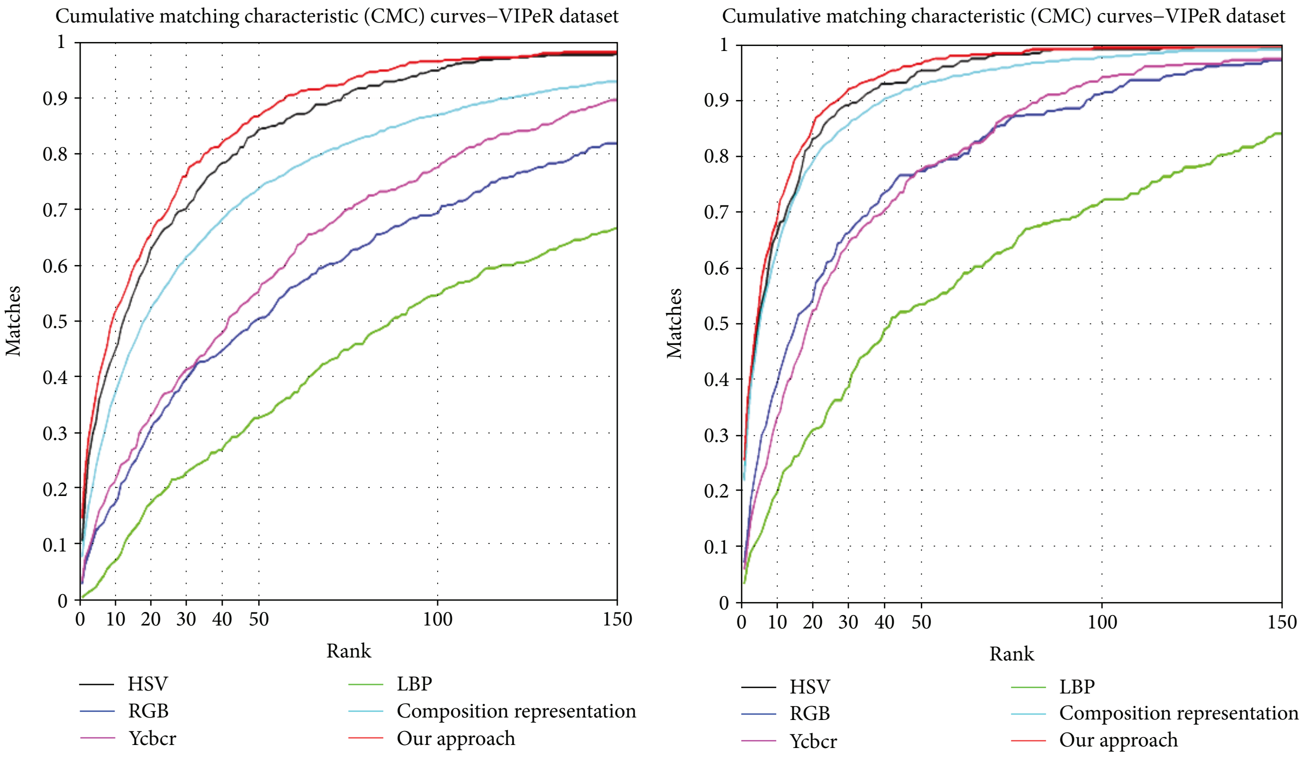
3.3. Experimental Setups and Evaluations. The extensive experiments are designed to include two phases: (1) The first phase examines the performance on different feature representations on the two datasets, which includes HSV, RGB, Ycbcr, LBP, and a composite representation of concatenating the four feature representations mentioned above, to see if the use of each feature representation or composite feature representation is enough to achieve a satisfactory performance. (2) Following that, further experiments are carried out to facilitate comparative testing with the state-of-the-art reidentification approaches including feature-based methods and metric learning-based methods on the two datasets. These approaches including LFDA [26], MFA [27], RS-KISS [42], RDC [61], Adaboost [62], Bhat [61], PLS [60], and Xing's [63] have also been reported in [42, 61].

Given the fact that the average cumulative match characteristic (CMC) [41] is commonly adopted by many existing reidentification approaches in the published literature, we also evaluate the performance of all the compared approaches using this criterion. Indeed, it treats person reidentification as a ranking problem. By providing a set of query images, the images in the target gallery are ranked according to their similarities to the query image. CMC curves measures the probability of a correct match. As the gallery size increases, it normally becomes more difficult to find the correct match and CMC curves become lower.

3.4. Experiment Results. Initially, we carried out the simulation tests on the VIPeR dataset. Such dataset is normally considered as the standard benchmark for a single-shot reidentification task. First, the images of P subjects are selected as the training set, while the rest are used as the testing set. In this experiment, we set $P = 100$ and $P = 316$, respectively, for training, where similar pairs are obtained from intraperson images of P subjects, and dissimilar pairs are obtained by randomly selecting interperson images from P subjects. Then, both similar pairs and dissimilar pairs are used to learn the RS-KISS metric on different feature spaces. Once the RS-KISS metric is learned, we are able to estimate weights of different feature representations by (10). Finally, the testing set is separated into the probe set and gallery set. The probe set consists of several single images from different subjects, while other single images of the same subjects are included in the gallery set. For the matching process, we select a single image from the probe set, and match it with all images from the gallery set based on the RS-KISS metric. This process is repeated for all images of the probe set. Following the same experimental protocol reported in [42], the average performance in the form of CMC curves over 10 different constructions of the probe set and gallery set is presented in Figure 4. To further investigate the effectiveness of the proposed approach, we further compared our approach with other popular person reidentification approaches. The results in terms of rank score are shown in the Table 1.

As shown in Figure 4, $P = 100$ samples are selected as the training data in Figure 4(a). While in Figure 4(b)$P = 316$ samples are selected as the training data. In each subfigure, the x-axis represents the rank score and y-axis represents the matching rate. The top 150 matching results are displayed in the figure. It is observed from Figure 4 that there is no single feature that can always achieve the best performance on the target dataset. Although composite representation is relatively better than single representation, its performance could not be guaranteed. In contrast, via weighted ensemble learning, our approach of combining four matching results obtained from four different feature representations always outperforms others.

Furthermore, we conduct one additional experiment on the VIPeR dataset to compare our approach with six other state-of-the-art person reidentification techniques mentioned in Section 3.3. For this experiment, we also set $P = 100$ and $P = 316$, respectively, for training, and report the averaged matching rates of 10 runs corresponding to rank scores = 1, 10, 25, and 50. As shown in Table 1, the results obtained by RS-KISS on composite representation is generally better than five competitive approaches including RDC, Adaboost, Bhat, PLS, and Xing's. On other hand, our approach further boosts the RS-KISS-based matching model, and achieves the best result by introducing the proposed weighted ensemble learning with multiple feature representations.



(a)

(b)

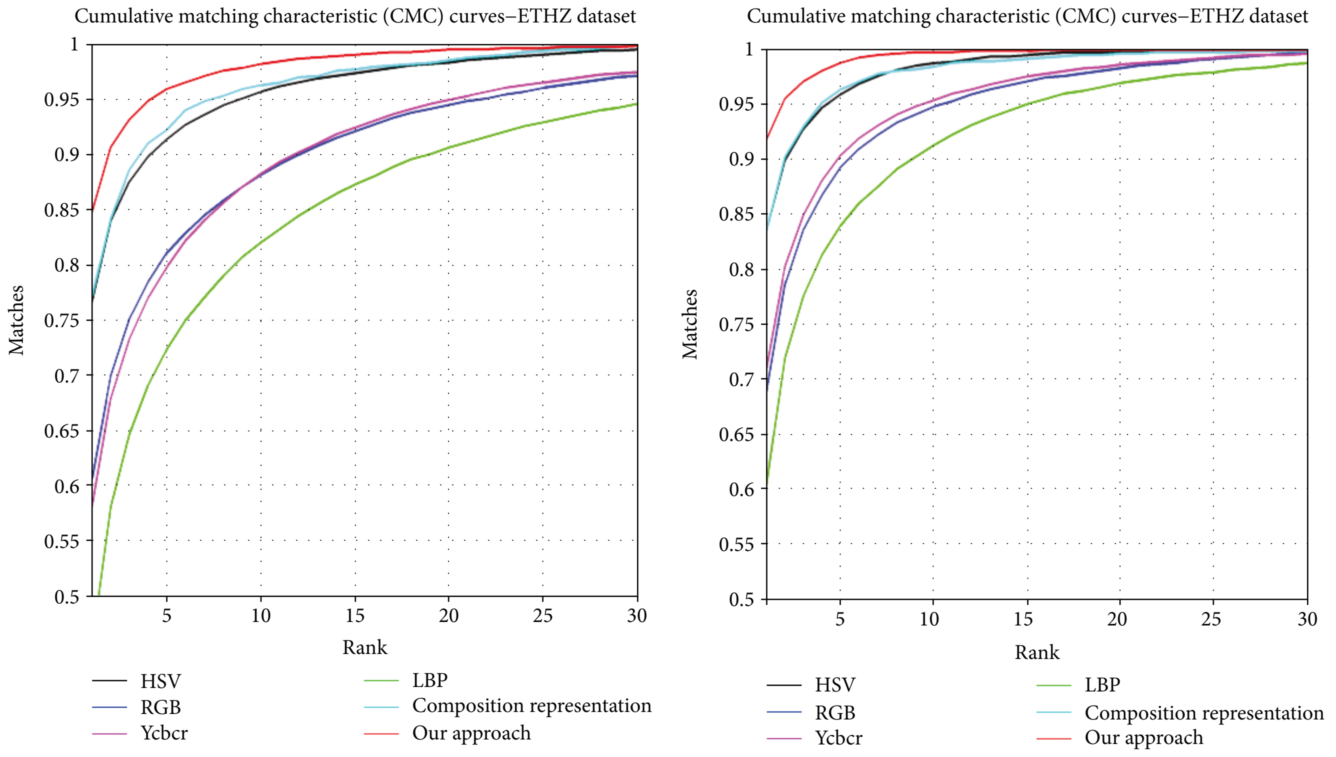
Figure 4: Performance on the VIPeR dataset in terms of CMC curves (top 150 ranking positions). $P = 100$ samples are selected as the training data in (a). $P = 316$ samples are selected as the training data in (b).

Table 1: Matching rates of different approaches on the VIPeR dataset.

Rank	$P = 100$				$P = 316$			
	1	10	25	50	1	10	25	50
Our approach	**0.147**	**0.513**	**0.701**	**0.867**	**0.285**	**0.763**	**0.915**	**0.987**
RS-KISS	0.098	0.405	0.608	0.765	0.245	0.666	0.847	0.930
RDC	0.091	0.344	0.535	0.697	0.157	0.539	0.752	0.879
LFDA	0.101	0.388	0.593	0.766	0.202	0.632	0.826	0.928
MFA	0.100	0.391	0.596	0.769	0.201	0.655	0.843	0.938
Adaboost	0.042	0.020	0.350	0.503	0.082	0.366	0.582	0.909
Bhat	0.038	0.124	0.203	0.295	0.047	0.166	0.266	0.402
PLS	0.023	0.082	0.142	0.232	0.27	0.109	0.204	0.329
Xing's	0.036	0.121	0.203	0.295	0.047	0.166	0.266	0.415

Next, we carry out the experiments on the ETHZ dataset. Such dataset provides a more realistic scenario of a multishot person reidentification task. For every target subject, it collects several images taken with a moving camera in different street scenes. All of the images of one person are obtained by the same camera with less viewpoint variation. In this part of the experiment, all sample images of $P = 76$ and $P = 106$ samples are selected as the training data, while the rest is used for testing. Following the same experimental setup on the VIPeR dataset, we also generate a set of similar and dissimilar pairs for the training process where the RS-KISS metric is learned and weights of different feature representations are estimated. Then, we randomly select one sample from every subject included in the testing set for the probe set and the rest for the gallery set. Both the probe and gallery sets are then used for testing.

The CMC curves of setting $P = 76$ and $P = 106$ are shown at Figures 5(a) and 5(b), respectively. Due to the fact that there are several images of a person in the gallery set on the ETHZ dataset, only the top 30 ranking positions are selected and shown in the figure. As a result, once again our approach achieves the best result when comparing either single feature representations or composite representation. Table 2 reports the averaged matching rates of 10 runs on the top 1, 5, 10, and 20 ranks for various person reidentification approaches. Of these, our approach also has the best performance.



(a) (b)

Figure 5: Performance on the ETHZ dataset in terms of CMC curves (top 30 ranking positions). $P = 76$ samples are selected as the training data in (a). $P = 106$ samples are selected as the training data in (b).

Table 2: Matching rates of different approaches on the ETHZ dataset.

Rank	$P = 76$				$P = 106$			
	1	5	10	20	1	5	10	20
Our approach	**0.852**	**0.958**	**0.981**	**0.994**	**0.918**	**0.987**	**0.997**	**0.999**
RS-KISS	0.770	0.921	0.962	0.985	0.835	0.963	0.984	0.996
RDC	0.690	0.858	0.922	0.969	0.727	0.901	0.956	0.988
LFDA	0.725	0.897	0.949	0.981	0.761	0.912	0.965	0.993
MFA	0.672	0.860	0.922	0.970	0.721	0.897	0.951	0.990
Adaboost	0.656	0.840	0.905	0.956	0.692	0.878	0.935	0.980
Bhat	0.555	0.761	0.840	0.906	0.610	0.809	0.878	0.941
PLS	0.483	0.694	0.780	0.868	0.546	0.751	0.833	0.924
Xing's	0.544	0.752	0.833	0.904	0.608	0.803	0.874	0.936

By close observation on both Figures 4 and 5, we can further release that our approach achieves much more performance gain on the ETHZ dataset than on the VIPeR dataset in comparison with second best one. On Tables 1 and 2, we are also able to observe that the performance gain at the top one rank is also much higher on the ETHZ dataset than on the VIPeR. Although few papers have published the matching results on multishot person reidentification, such results strongly indicate that our approach not only performs well on a single-shot person reidentification task but also achieve outstanding performance on multishot person reidentification.

4. Discussion

As the first and most straightforward visual feature, color plays an important role for the person reidentification task, but changes in brightness may lead to instability in such features. On the other hand, texture and structure features give us information on the structural arrangement of surfaces and objects in the image, and take much more effect when the appearance contains distinct partial patterns. Conceptually, different kinds of feature representations obtained from different aspects, for example, color versus texture, and on different scales, for example, local versus global, as well as fine versus coarse, always tend to be complementary in improving reidentification accuracy. Therefore, it is nontrivial to fuse various feature representations for robust person reidentification. In this work, we only select four simple feature representations for demonstration purposes. It will be an interesting future research to systematically explore the complementary nature between different feature representations. Even so, our experiment results still show that the matching model on the ensemble of four selected feature representations significantly outperforms the one with a single representation; moreover, a single matching model working on a composite representation formed by concatenating four selected feature representations together is often inferior to an ensemble of multiple matching models on different representations. Therefore, we strongly believe that the proposed weighted ensemble learning model is more effective and efficient than a single learning model on the composite representation of a much higher dimension.

Ensemble learning provides an underpinning yet enabling technique of combining multiple matching obtained from different feature representations for reidentification tasks. But a fundamental weakness in ensemble learning is that different base learners are normally treated equally during reconciliation. In this work, the equation derived in (9) reveals that the performance of ensemble learning depends on both the quality of multiple matching results and a weighted ensemble scheme. The first term of (9) indicates that the quality of matching results is essentially determined by the discriminative power of corresponding feature representations, which are quantified by a set of weights. Theoretically, the bigger value of the weights should be assigned to the better matching results with a greater discriminative power of the feature representation, and vice versa. Then, an optimal result can be finally obtained by combining these matching results via a weighted ensemble scheme. Our previous works [51, 52] also confirm that the weighted ensemble scheme normally outperforms the averaged ensemble in terms of both effectiveness and efficiency. As a wrapper learning technique, ensemble learning has been widely used in many applications, and aims to boost the performance of a learning-based system by combining multiple base learners into an optimal consensus solution.

Our experiment results also show that the performances of the tested approaches are improved by increasing the size of the training set. In other words, the available amount of the training image pairs crucially decide whether a distance metric is sufficiently learned for person reidentification. However, in practice, it is quite expensive to obtain the desired amount of training set with label information. Therefore, the RS-KISS metric is adopted in our approach due to its outstanding ability of dealing with limited training data. In fact, RS-KISS intends to improve the original version of KISS when the size of the training set is small, because the learning process of the covariance matrix in KISS is always biased on a small-size training set. Although RS-KISS performs comparably to KISS when the training sample set is large enough, it normally incurs a higher computational burden due to the composite representation of a much higher dimension. In contrast, our approach provides an optimal solution by constructing a weighted ensemble of multiple matching results obtained on different feature representations, which significantly improves the performance of reidentification.

5. Conclusions

In this paper, we present a novel person reidentification technique by proposing weighted ensemble learning with different feature representations. In our approach, we adopt the RS-KISS metric in the matching process which keeps its excellence of dealing with an insufficient training set. Initially, the RS-KISS metric is correspondingly learned on four selected feature spaces of the training set. Then, a set of weights are estimated to access the importance of different feature representations according to its discriminative power. Finally, the testing stage is carried out by combining the multiple matching results obtained from different feature representations into an optimal one via the weighted ensemble scheme. As a generic framework, our weighted ensemble module generally allows any feature representation to be incorporated directly. In our experiments, results show that our approach is very competitive in comparison with several state-of-the-art approaches, and thus provides a promising technique for person reidentification.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the financial support given by the Natural Science Foundation of China (NSFC) under Grant nos. 61402397, 61663046, and 61572486, Applied Basic Research Foundation of Yunnan Province under Grant nos. 2016FB104 and 2016FB105, Yunnan Science and Technology Fund (2017FA034), Yunnan Province Young Academic and Technical Leader Candidate Program (2018HB027, 2017HB005), Yunnan Provincial E-Business Entrepreneur Innovation Interactive Space (2017DS012), Kunming Key Laboratory of E-Business and Alternative Finance (2017-1A-14684, KGF2018), and Yunnan Provincial E-Business Innovation and Entrepreneurship Key Laboratory of Colleges and Universities (YES2014).

References

[1] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, “Person re-identification by descriptive and discriminative classification,” in *Image Analysis*, pp. 91–102, Springer, 2011.

[2] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary, “Person re-identification by support vector ranking,” in *Procedings of the British Machine Vision Conference*, pp. 21.1–21.11, BMVA Press, 2010.

[3] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, “Person re-identification by symmetry-driven

accumulation of local features," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2360–2367, San Francisco, CA, USA, June 2010.

[4] S. Bak, E. Corvee, F. Brémond, and M. Thonnat, "Person re-identification using spatial covariance regions of human body parts," in *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 435–440, Boston, MA, USA, 2010.

[5] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by fisher vectors for person re-identification," in *Computer Vision—-ECCV 2012. Workshops and Demonstrations*, pp. 413–422, Springer, 2012.

[6] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, Rio de Janeiro, Brazil, October 2007.

[7] M. Dikmen, E. Akbas, T. S. Huang, and N. Ahuja, "Pedestrian recognition with a learned metric," in *Computer Vision–ACCV 2010*, pp. 501–512, Springer, 2010.

[8] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[9] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biological Cybernetics*, vol. 61, no. 2, 1989.

[10] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image and Vision Computing*, vol. 32, no. 4, pp. 270–286, 2014.

[11] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: a SIFT descriptor with color invariant characteristics," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 1978–1983, New York, NY, USA, June 2006.

[12] Y. Zhang and S. Li, "Gabor-LBP based region covariance descriptor for person re-identification," in *2011 Sixth International Conference on Image and Graphics*, pp. 368–371, Hefei, Anhui, China, August 2011.

[13] W. Ayedi, H. Snoussi, and M. Abid, "A fast multi-scale covariance descriptor for object re-identification," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1902–1907, 2012.

[14] C. Liu, S. Gong, C. C. Loy, and X. Lin, "Person re-identification: what features are important?," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*, pp. 391–401, Springer, 2012.

[15] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3586–3593, Portland, OR, USA, June 2013.

[16] R. Layne, T. M. Hospedales, S. Gong, and Q. Mary, "Person re-identification by attributes," in *Procedings of the British Machine Vision Conference*, BMVA Press, 2012.

[17] R. Shang, W. Wang, R. Stolkin, and L. Jiao, "Subspace learning-based graph regularized feature selection," *Knowledge-Based Systems*, vol. 112, pp. 152–165, 2016.

[18] Y. Wang, T. Shen, G. Yuan, J. Bian, and X. Fu, "Appearance-based gaze estimation using deep features and random forest regression," *Knowledge-Based Systems*, vol. 110, pp. 293–301, 2016.

[19] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *2014 22nd International Conference on Pattern Recognition*, pp. 34–39, Stockholm, Sweden, August 2014.

[20] W. Li, R. Zhao, T. Xiao, and X. Wang, "DeepReID: deep filter pairing neural network for person re-identification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 152–159, Columbus, OH, USA, June 2014.

[21] W. Chen, X. Chen, J. Zhang, and K. Huang, "A multi-task deep network for person re-identification," in *Conference on Artificial Intelligence*, AAAI, 2017.

[22] F. Wang, W. Zuo, L. Lin, D. Zhang, and L. Zhang, "Joint learning of single-image and cross-image representations for person re-identification," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1288–1296, Las Vegas, NV, USA, June 2016.

[23] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2197–2206, Boston, MA, USA, June 2015.

[24] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.

[25] M. Loog, R. P. W. Duin, and R. Haeb-Umbach, "Multiclass linear dimension reduction by weighted pairwise fisher criteria," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 7, pp. 762–766, 2001.

[26] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local fisher discriminant analysis for pedestrian re-identification," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3318–3325, Portland, OR, USA, June 2013.

[27] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.

[28] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in *2011 IEEE conference on Computer vision and pattern recognition*, pp. 649–656, Colorado Springs, CO, USA, June 2011.

[29] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch, "Learning implicit transfer for person re-identification," in *Computer Vision—ECCV 2012. Workshops and Demonstrations*, pp. 381–390, Springer, 2012.

[30] L. Bazzani, M. Cristani, and V. Murino, "Symmetry-driven accumulation of local features for human characterization and re-identification," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 130–144, 2013.

[31] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification," in *Proceedings of the British Machine Vision Conference*, pp. 68.1–68.11, Dundee, UK, 2011.

[32] B. Ma, Y. Su, and F. Jurie, "BICov: a novel image representation for person re-identification and face verification," in *Proceedings of the British Machine Vision Conference*, R. Bowden, J. Collomosse, and K. Mikolajczyk, Eds., pp. 57.1–57.11, BMVA Press, 2012.

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

[34] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *The Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

[35] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *ICML '07 Proceedings of the 24th International Conference on Machine Learning*, pp. 209–216, New York, NY, USA, June 2007.

[36] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2288–2295, Providence, RI, USA, June 2012.

[37] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in *2009 IEEE 12th International Conference on Computer Vision*, pp. 498–505, Kyoto, Japan, 2009.

[38] A. Mignon and F. Jurie, "PCCA: a new approach for distance learning from sparse pairwise constraints," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2666–2672, Providence, RI, USA, June 2012.

[39] R. Satta, G. Fumera, F. Roli, M. Cristani, and V. Murino, "A multiple component matching framework for person re-identification," in *Image Analysis and Processing—ICIAP 2011*, vol. 6979 of ICIAP 2011. Lecture Notes in Computer Science, , pp. 140–149, Springer, 2011.

[40] R. Satta, G. Fumera, and F. Roli, "Exploiting dissimilarity representations for person re-identification," in *Similarity-Based Pattern Recognition*, pp. 275–289, Springer, 2011.

[41] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, Citeseer, Rio de Janeiro, 2007.

[42] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li, "Person re-identification by regularized smoothing KISS metric learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 10, pp. 1675–1685, 2013.

[43] M. S. Roy and S. K. Bandyophadyay, "Face detection using a hybrid approach that combines HSV and RGB," *International Journal of Computer Science and Mobile Computing*, vol. 2, no. 3, pp. 127–136, 2013.

[44] M. W. Schwarz, W. B. Cowan, and J. C. Beatty, "An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models," *ACM Transactions on Graphics*, vol. 6, no. 2, pp. 123–158, 1987.

[45] D. Chai and A. Bouzerdoum, "A Bayesian approach to skin color classification in YCbCr color space," in *2000 TENCON Proceedings. Intelligent Systems and Technologies for the New Millennium (Cat. No. 00CH37119)*, pp. 421–424, Kuala Lumpur, Malaysia, 2000.

[46] S. L. Phung, A. Bouzerdoum, and D. Chai, "A novel skin color model in YCbCr color space and its application to human face detection," in *Proceedings International Conference on Image Processing*, pp. I-289–I-292, Rochester, NY, USA, September 2002.

[47] F. Kimura, K. Takashina, S. Tsuruoka, and Y. Miyake, "Modified quadratic discriminant functions and the application to Chinese character recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 1, pp. 149–153, 1987.

[48] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, vol. 84, no. 405, pp. 165–175, 1989.

[49] Z. Q. John Lu, "The elements of statistical learning: data mining, inference, and prediction," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 173, no. 3, pp. 693-694, 2010.

[50] Y. Yang and J. Jiang, "Adaptive bi-weighting toward automatic initialization and model selection for HMM-based hybrid meta-clustering ensembles," *IEEE Transactions on Cybernetics*, pp. 1–12, 2018.

[51] Y. Yang and J. Jiang, "Hybrid sampling-based clustering ensemble with global and local constitutions," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 5, pp. 952–965, 2016.

[52] Y. Yang and K. Chen, "Temporal data clustering via weighted clustering ensemble with different representations," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 2, pp. 307–320, 2011.

[53] Y. Yang and K. Chen, "Time series clustering via RPCL network ensemble with different representations," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 2, pp. 190–199, 2011.

[54] Y. Yang and J. Jiang, "HMM-based hybrid meta-clustering ensemble for temporal data," *Knowledge-Based Systems*, vol. 56, pp. 299–310, 2014.

[55] Y. Yang and J. Jiang, "Bi-weighted ensemble via HMM-based approaches for temporal data clustering," *Pattern Recognition*, vol. 76, pp. 391–403, 2018.

[56] S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Learning to rank in person re-identification with metric ensembles," in *2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1846–1855, Boston, MA, USA, June 2015.

[57] B. Seijo-Pardo, I. Porto-Díaz, V. Bolón-Canedo, and A. Alonso-Betanzos, "Ensemble feature selection: homogeneous and heterogeneous approaches," *Knowledge-Based Systems*, vol. 118, pp. 124–139, 2017.

[58] M. S. Akhtar, D. Gupta, A. Ekbal, and P. Bhattacharyya, "Feature selection and ensemble construction: a two-step method for aspect based sentiment analysis," *Knowledge-Based Systems*, vol. 125, pp. 116–135, 2017.

[59] A. Ess, B. Leibe, and L. Van Gool, "Depth and appearance for mobile scene analysis," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, Rio de Janeiro, Brazil, October 2007.

[60] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," in *2009 XXII Brazilian Symposium on Computer Graphics and Image Processing*, pp. 322–329, Rio de Janiero, Brazil, October 2009.

[61] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 653–668, 2013.

[62] S. Bak, E. Corvee, F. Brémond, and M. Thonnat, "Person re-identification using Haar-based and DCD-based signature," in *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pp. 1–8, Boston, MA, USA, 2010.

[63] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning with application to clustering with side-information," *Advances in Neural Information Processing Systems*, vol. 15, pp. 505–512, 2003.