

Performance Evaluation and Scheme Selection of Person Re-Identification Algorithms in Video Surveillance

Yu Lu, School of Information Science and Engineering, Shandong University, Qingdao, China

Lingchen Gu, School of Information Science and Engineering, Shandong University, Qingdao, China

Ju Liu, School of Information Science and Engineering, Shandong University, Qingdao, China

Peng Lan, School of Information Science and Engineering, Shandong Agricultural University, Tai'an, China

ABSTRACT

With the increasing number of camera networks deployed in public places, intelligent video processing has become a key technology for video surveillance. In order to alleviate the workload of the tracers in the artificial tracking video, person re-identification (re-id) can match a large number of pedestrian images to obtain the location of same person at different time in surveillance. This article focuses on the comparison of different classic distance metric learning methods so as to select optimum person re-identification scheme with excellent performance. The authors compare four algorithms matching Local Maximal Occurrence (LOMO) feature representation on three common databases and obtains a criterion to choose algorithms for different datasets. The selection of re-identification algorithms can simplify the video investigation process according to the size and number of person images. In the end, they propose an improved metric learning based on one of algorithms and get improved results. The re-id is useful and efficient in works such as the criminal investigators etc.

KEYWORDS

Algorithms Comparison, Distance Metric, Feature Representation, Person Re-identification

1. INTRODUCTION

In recent years, with the rapid development of image acquisition and large-scale data storage technologies, large scale camera-nets have produced rich video surveillance information. Video analysis technology can realize the recognition and tracking of designated pedestrians in video, which greatly releases the pressure of manual viewing, and improves the effectiveness of monitoring system. Person re-identification is to identify the target persons in the video sequences from the non-overlapping camera view. Re-identification (re-id) tries to solve the problem of “where the target person was presented in the past” or “where the target person was captured in the surveillance in the future”. Although in a controllable environment, it is a relatively mature technology to recognize persons based on human faces and other biological features. However, because the environment of monitoring video is complicated and flexible, the resolution of the extracted person images is low. So it is difficult to obtain robust face features. Therefore, person re-id often uses the appearance features of the clothes and some significant things carried with target persons.

DOI: 10.4018/IJDCF.2019100104

This article, originally published under IGI Global's copyright on October 1, 2019 will proceed with publication as an Open Access article starting on February 2, 2021 in the gold Open Access journal, International Journal of Digital Crime and Forensics (converted to gold Open Access January 1, 2021), and will be distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

The re-id problem can be summarized as follows: there are several cameras from non-overlapping views. We obtain a number of person images by preprocessing technology. The images in cameras 2, 3... defined as probe and images in camera 1 defined as gallery. The aim of re-identification is to find out special images containing the target persons which are appointed in probe. The sketch map of re-id as shown in Figure 1.

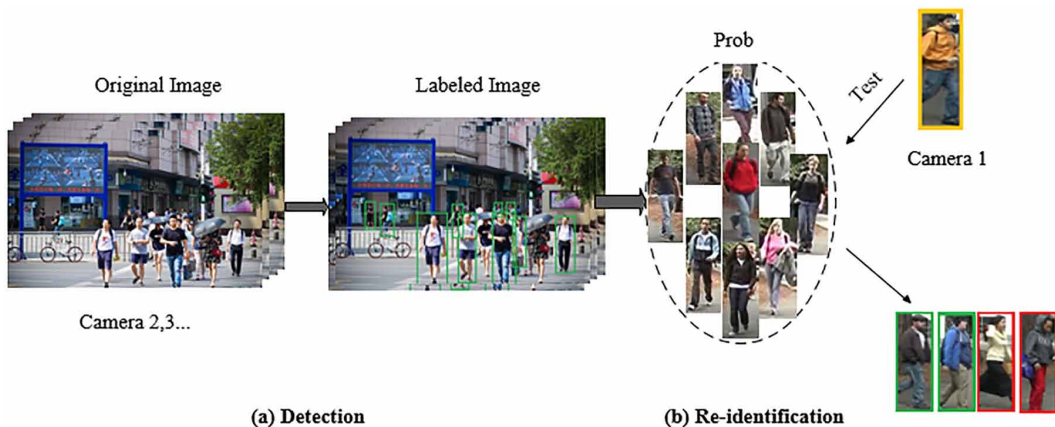
In the application of the real scenes, the re-id faces some essential problems:

1. The differences of views result in the misalignment of person appearance. In fact, the in-stallation of cameras is usually fixed in practical applications, but the visual angle under various cameras is often different. This diversity can even result in the problem that some features which can be seen in camera A will not be found under camera B. The deviation of appearance position will cause the same person could not be discovered in different videos. In existing public databases, more than 90% of person images exist this difference of perspective.
2. The traditional distance function does not consider the sample characteristics and this problem lead to a weak discriminant. The distance functions such as Euclidean distance, cosine distance and correlation distance have bad discriminant ability because they do not take the distribution characteristics of samples into account. According to the experimental statistics, the recognition rate of the measure function based on sample distribution learning in VIPeR database is about 32% higher than the traditional distance ones.

The key technologies of person re-id include feature representation, feature transformation, distance measurement and sorting. These days, more scientists put their concentrate on two technologies: feature representation and distance learning. Typical algorithms of person re-id now can be ranked as follows:

1. Design a robust feature representation. In order to design a stable and discriminant feature representation, the dimensions of feature vectors have been increasingly high, from thou-sands to tens of thousands.
2. Learn a good distance metric. Learning a robust and discriminant metric function or cross-view subspace can improve re-id performances.

Figure 1. Person re-id sketch map



Nevertheless, due to the differences in visual angle and illumination of different monitoring videos, the extracted features of persons change sharply, which may lead to a more similar appearance of distinct persons than the same person in a number of videos. Hence, a stable feature representation is the first step to achieve person re-id. Then, academic circle often calculates feature vectors of two-person images to match them, and require the distance between same persons must less than different ones. Above all, we can conclude that feature representation and metric learning are two key technologies in person re-id.

In this paper, we choose the Local Maximal Occurrence (LOMO) feature which was generated by Liao in 2015 (Liao et al., 2015). The LOMO had achieved higher performance by traditional labeled ways. In terms of metric learning, we use LOMO to match four learning algorithms on three common datasets to complete re-id, and compare the performances and feature of dataset' images to conclude which learning algorithm suitable for target datasets. In addition, we propose an improved metric learning method and get better results.

2. RELATED WORK

The existing feature representation methods have extended from original color features to texture and semantics. However, color features are still the most widely used visual features in image retrieval. The color is often closely related to objects or scenes contained in the image. In addition, compared with other visual features, color features have less dependence on the size, direction and angle of the image itself, and thus have higher robustness. Color histogram is widely used to represent feature vectors in person re-id. Histogram describes the proportion of different colors in the whole image. Although most digital images are expressed in RGB color space, the spatial results do not conform to people's subjective judgment of color. Therefore, some people propose histograms based on HSV space, LUV space and LAB space, in which the HSV space is the most common color space of the histogram (Yang et al., 2014), and its three components represent hue, saturation and value respectively.

Texture is a kind of visual feature reflecting the homogeneous phenomenon in the image. It embodies the structure and arrangement properties of surface structure with slow change or periodic change. Different with the color features, the texture is represented by the gray distribution of pixels and the surrounding spatial domain. i.e., the local texture information. While texture features embody the properties of global features, it also describes the surface properties of the scene corresponding to the image or image region. However, as the texture is only a feature of an object surface, it cannot completely reflect the essential attribute of the object, so it is impossible to obtain the high-level image content only by using the texture features. Unlike color features, texture features are not based on pixel features, but need to be statistically computed in areas containing multiple pixels. In pattern matching, this regional feature has great superiority, and cannot be unsuccessful because of local deviation.

Texture feature is an effective method when retrieving texture images with large differences in thickness and density. But when there is little difference among the easily distinguishable information between the texture and the thickness, the usual texture features can hardly accurately reflect the difference between the different textures of the human visual sense. For example, the reflection of water, the reflection of smooth metal surfaces, and so on, all affect the texture. Because these are not the characteristics of the object itself, when texture information is applied to the retrieval, sometimes these false textures will cause "misleading" to the retrieval. In addition, texture features have excellent anti-noise performance, but are also vulnerable to illumination changes.

Above all, we use LOMO feature representation matching four metric learning methods to achieve re-id. LOMO feature analysis the local features of horizontal events and maximizes event enhancement stability, thus resisting visual changes. In addition, this method also uses the deformation of Retinex, which can effectively handle the illumination change problem in re-identification.

Dealing with Illumination Variations:

In academic world, color is usually regarded as an important feature to describe person images. However, different cameras are under different lighting conditions, and the camera models are varied. Even if the human eye feels the same color to the same persons, the color perception may be different from different camera environments.

Liao firstly used Retinex algorithm to preprocess person images. From the perspective of person illumination and color, Retinex generates color images consistent with human eye observation scenes. The restored image contains rich color information, especially for the details of the shadow area.

The multiscale Retinex algorithm (Jobson et al., 1997; Land et al., 1971) is used in LOMO feature. Multiscale Retinex fusion, small-scale Retinex dynamic range compression and large scale Retinex simultaneously perform color reproduction. Therefore, this algorithm is often used to deal with the stability of color persistence and dynamic range of automatic compression, to achieve good near-human visual perception. In LOMO feature, two scales of Retinex are used, the parameters are 5 and 20 respectively. It is obvious that the processed images have better consistency in brightness and color, which makes it easier for person re-id than before using the original images. Then, HSV color histogram is extracted from the Retinex images as part of the feature information.

Liao also applied the Scale Invariant Local Ternary Pattern (SILTP) descriptor for illumination invariant texture description (Liao et al., 2010). In fact, LBP has good stability under monotonic gray transformation, but its robustness to image noise is poor. SILTP is an improved operator over the well-known Local Binary Pattern (LBP). It introduces a scale invariant local comparison tolerance and achieves robustness to image noise.

Dealing with Viewpoint Change:

Persons from different cameras usually appear at different angles. For example, a person in the front view of a camera may appear in the rear view of another camera. Therefore, matching persons from different perspectives is also a difficulty in re-id. In order to solve this problem, Liao proposes to divide the person image into six horizontal fringes and compute a histogram in each fringe. However, this method may also lose the spatial detail information in the stripes and affect its identification ability.

Liao used sliding windows to describe the local details of person images. They used a sub-window size of 10*10, with an overlapping step of 5 pixels to locate local patches in 128*48 images. Within each sub-window, he extracted two scales of SILTP histograms ($SILTP_{4,3}^{0.3}$, $SILTP_{4,5}^{0.3}$), and an 8*8*8-bin joint HSV histogram. Each histogram bin represents the occurrence probability of one pattern in a sub-window. To solve the problem of view changes, Liao examines all the sub-windows at the same horizontal position (Prosser et al., 2010) and maximizes the possibility of local occurrence of each pattern. The generated histogram can achieve stability under the change of angle and capture the local characteristics of humans.

Considering the multiscale problem, they also built a three-scale pyramid representation for the purpose of multi-scale information, which down-samples the original 128*48 image by 2*2 local average pooling operations. His final descriptor had (8*8*8 color bins + 34 *2 SILTP bins) * (24+11+5 horizontal groups) = 26,960 dimensions.

3. METRIC LEARNING

In this session, we will first briefly summarize the learning principles of four distance metric algorithms (Zheng et al., 2011).

3.1. Cross-view Quadratic Distance Algorithm (XQDA)

Under the precondition of zero mean Gaussian distribution, Liao et al., (2015) extended Bayesian Face and KISSME to compare covariance matrix \sum_I and \sum_E of intra-class Ω_I and inter-class Ω_E (Moghaddam et al., 2002), respectively. Considering the original feature dimensions d is large,

and a low dimensional space is preferred for classification. XQDA learns a distance function in the r dimensional subspace for the cross-view similarity measure. Suppose there is a cross-view training set $\{X, Z\}$, where $X = (x_1, x_2, \dots, x_n) \in R^{d \times n}$ contains n samples in a d dimensional space from one view, $Z = (z_1, z_2, \dots, z_m) \in R^{d \times m}$ contains m samples in the same d dimensional space but from the other view. The cross-view matching issue arises from many applications, such as face recognition (Liao et al., 2009) and viewpoint invariant person re-id. It is important to note that Z is the same with X in the single-view matching scenario. The distance between examples x_i and z_j is defined as form of Mahalanobis distance

$$d(x_i, z_j) = (x_i - z_j)^T \left(\sum_I^{-1} - \sum_E^{-1} \right) (x_i - z_j) \quad (1)$$

The kernel matrix is $M = \sum_I^{-1} - \sum_E^{-1}$. However, it is difficult to solve this kernel matrix directly because it contains two inverse matrices. So transferring the M to the subspace by reducing the feature vectors dimension, the kernel matrix reshaped as

$$M(W) = W \left(\sum_I^{-1} - \sum_E^{-1} \right) W^T \quad (2)$$

In order to avoid the inconvenience of learning two inverse matrices, Liao optimizes the ratio of two covariance $\frac{\sigma_E(w)}{\sigma_I(w)}$ to obtain the optimal solution

$$J(w) = \frac{w^T \sum_E w}{w^T \sum_I w} \quad (3)$$

where $\sigma_I(w) = w^T \sum_I w$ and $\sigma_E(w) = w^T \sum_E w$.

The maximization of $J(w)$ is equivalent to $\max_w w^T \sum_E w$, s.t. $w^T \sum_I w = 1$.

4. MLAPG

Matrix learning plays a very important role in image matching of person re-id. However, if the positive semidefinite (PSD) constraint is applied on the basis of existing distance functions, a large number of calculations will be generated. Conversely, the distance function of free learning will produce potential noise. In this section, we introduce that PSD constraint can provide an effective regularization method to smooth the solution of metric learning, so the learned metric is more robust than without the PSD constraint.

Another key problem is that there is a serious imbalance between the number of positive and negative samples in the learning process of re-id. In particular, some databases for person re-id (such as VIPeR and QMUL_Grid) are small, resulting in a limited number of positive samples. Therefore, the learning process is easily controlled by a large number of negative sample pairs, which leads to poor learning results. Moreover, the problem of imbalance is even more affected than above PSD constraint.

In this section, we learn a logical metric learning method of a PSD constrained asymmetric sample weight strategy in terms of distance measurement. In addition, this method finds the global minimum solution of the objective function by accelerating the near-end gradient method (APG).

Suppose there is cross-view training set $\{x, z, y\}$, where $x = (x_1, x_2, \dots, x_n) \in R^{d \times n}$ contains samples in a d dimensional space from one view, $z = (z_1, z_2, \dots, z_m) \in R^{d \times m}$ contains m samples in the same d dimensional space but from another view, and $y \in R^{n \times m}$ is the matching label between x and z . Like XQDA, MLAPG (Liao et al., 2015) also used Mahalanobis distance as Equation (4) (Xing et al., 2002) to solve image matching problem

$$D_M^2(x, z) = \|x - z\|_M^2 = (x - z)^T M (x - z) \quad (4)$$

But different from the above method, Liao applied PSD (Semi-positive Definite Matrix) constraint (Kostinger et al., 2012), and transferred the solution of the distance function to the minimization problem of a smooth convex loss function $f_M(x, z) = \log\left(1 + e^{y(D_M^2(x, z) - \mu)}\right)$ (Zheng & Xiang 2013) in logistical form. Hence, the expression (Nesterov 2004) in the whole dataset is

$$F(M) = \sum_{i=1}^n \sum_{j=1}^m w_{ij} f_M(x_i, z_j) \quad (5)$$

It is worth mentioning that Liao proposed an asymmetric sample weighting strategy, where $w_{ij} = \frac{1}{N_{pos}}$ when $y_{ij} = 1$ and $w_{ij} = \frac{1}{N_{neg}}$ when $y_{ij} = -1$ (N_{pos} and N_{neg} are the number of positive and negative sample pairs respectively). At the end of computation, the cross-view logistic learning function is simplified as

$$\min_M F(M) \quad , \quad s.t. M \geq 0 \quad (6)$$

4.1. Discriminative Null Space (DNS)

Zhang et al. (Zhang et al., 2016) proposed that project the same person to one point in null space.

Null Foley-Sammon transform:

Null Foley-Sammon transform (NFST) is in order to learn a discriminative subspace in which the training data points in every C classes are collapsed to a single point. NFST also ensures that C points should not further collapse to a single point for the purpose of discriminative. The author aims to learn an optimal projection matrix W , so that each of its column w satisfies the following two conditions: $w^T S_w w = 0$, $w^T S_b w > 0$.

Hence, it guarantees the best separability of training data in the term of Fisher discriminant criterion. Zhang calls this linear projecting direction w as Null Projecting Direction (NPD) (Guo et al., 2006).

In conclusion, the key of learning the discriminative null space depends on solving an eigen-problem, which has a very efficient closed-form solution. Particularly, the whole optimization algorithm has no free parameter to tune.

Semi-supervised Learning:

When applied the fully-supervised method NFST to the problem of re-id, we use labelled training set learn the projection W . Then, the test data are projected into the same subspace and matched by computing the Euclidean distance between a query sample and a set of gallery samples.

In practical applications, the number of labeled training data is sparse, while the number of unlabeled data (person images under different points) is abundant that can alleviate the small sample size problem. The NFST algorithm is extended to the semi-supervised setting for this purpose in Figure 2.

4.2. Enhanced Self-Trained Learning (SSL)

Although many researchers has made much promising efforts in distance metric learning, most existing methods are developed in the fully-supervised setting, requiring access to a large amount of labeled training image pairs. It is impossible to obtain a large quantities of labeled data because labeling images is very costly. The main bottleneck for re-id under fully-supervised is the limited access to labeled training samples. It is prone to over-fitted to the labeled data when only a small number of labeled data are available. So it is important to design a solution that can utilize abundant unlabeled data (Chapelle & Zien 2009). Yang formulates person re-id as a subspace learning problem, which is similar with (An & Bhanu 2015; Pedagadi et al., 2013). They (Yang et al., 2017) design a self-trained (Zhu & Goldberg 2009) subspace learning approach for person re-id which effectively utilizes both labeled and unlabeled data to learn a discriminative subspace where person images across disjoint camera views can be matched easily.

Yang proposed the enhancing semi-supervised (Cai et al., 2007; Hoi et al., 2008) self-learned algorithm. The core thought is to use some of labeled image data X_l to train learning function. First, Yang learns an initial projection matrix using the available data only. Then he constructs pseudo pairwise relationships among unlabeled person images X_u using k-nearest (KNN) algorithm in this low-dimensional subspace Z_u using the initial projection. The low-dimensional representation is $Z_u = (U^0)^T X_u$. A KNN graph is usually constructed to model the relationship between nearby data nodes, where an edge will be placed between node i and node j using original feature representation.

Figure 2. Semi-supervised null space learning

Algorithm 1: Semi-supervised null space learning

Input: $X^l, X^u, P^0 = 0$

Output: The learned projection W .

- 1: Estimate W^0 using X^l ;
 - 2: $t = 0$;
 - 3: **while** not converged **do**
 - 4: project X^u through W^t to obtain $Y_{W^t}^u$
 - 5: build k-nn graph G with $Y_{W^t}^u$
 - 6: take top f percent to create the pseudo-classes P^{t+1}
 - 7: learn W^{t+1} with $X^l + P^{t+1}$
 - 8: $t = t + 1$
 - 9: **end while**
-

The pseudo pairwise relationships are encoded into a graph Laplacian regularization term which is further combined with a fully-supervised discriminative term to learn a new projection. Given the newly learned projection, Yang refines the relationships and relearned the discriminative projection with these updated pseudo pairwise relationships.

4.3. Our Proposed Method

In above section, we introduce the SSL which transforms the person re-id into discriminant projection matrix learning in subspace by iteratively processing labeled and unlabeled data. The advantage of this semi-supervised algorithm is to use only part of labeled information in the database to match unlabeled person images from disjoint camera views to pseudo pairwise relationships. Iterative learning is needed in the matching process to enhance the learning performance.

However, when implementing the SSL method on the database, Yang (Yang et al., 2017) does not consider the impact of labeled information utilization ratio on the learning effect during the iteration process. In Equation 4, the small regularization parameter η is the trade-off between the first labeled term and the second unlabeled term, but η is always equal to one during the whole experiments, and does not change with the iteration process.

Considering that η is the parameter of the unlabeled term, and with the increase of iterations, the relearning projection matrix constantly updates the pseudo pairwise relationships to make the matching results more accurate. Therefore, we increase the parameter η of the unlabeled term during iteration to enhance the re-id performance.

This small change of SSL has gained experimental performance improving. The iterative method is shown as in the following Figure 3.

5. EXPERIMENTAL SECTION

5.1. Experimental Environment and Data

All experiments in this paper performed under MATLAB 2015b circumstance. The parameters of computer are as follows in Table 1

We choose three common public datasets VIPeR, CUHK01 and CUHK03.

5.2. Experimental Results and Analysis

VIPeR: It is a dataset widely used as a training test in person re-id, which includes 632 persons with 1264 images, and the images were collected from two cameras in the outdoor environment. Each person has only one image under one camera and all images are 128*48 pixels. In VIPeR, person images are significantly different in background, light and angle of view, so there is great difference in people appearance. During experiment, we divide half persons into training set and the other half as testing set.

CUHK01: The CUHK01 dataset is a homemade person re-id database built by Chinese University Hong Kong. Unlike other databases, person images in CUHK01 have high resolution. There are 971 persons in these databases. Each person has two images under each shooting angle. If cam A takes the front view of one person, cam B will take the other view of this person. All images in CUHK01 are the same as 160*60. In the experiment, 485 persons are trained, while the remaining 486 are used for testing.

CUHK03: The CUHK03 database has 13164 images of 1360 persons, and it is also one of the most popular person re-id databases. CUHK03 had been collected for several months, involving a total of 6 cameras, averaging 4.8 images per view. Images in this database have practical significance because the problems such as pedestrian ectopic and lack of body parts. In the experiment, 1160 persons are used as training sets, and the remaining 200 persons are used as testing sets.

Figure 3. The proposed self-trained subspace learning approach

Algorithm 1 The proposed self-trained subspace learning approach

Input: The labeled training data X_l and its weight matrix

W^l ; the unlabeled training data X_u ; the parameter

η ; the maximal number of iterations T .

Initialize: $U^0 = \arg \min_U L(X_l, U, W^l)$; $t = 1$.

Iterative: $t = 1, 2, \dots, T$

1: Project X_u into a low-dimensional subspace through $Z_u^T = (U^{t-1})^T X_u$;

2: Build the pseudo pairwise relationships among the unlabeled data by constructing a KNN graph using Z_u^T ;

3: Encode the pseudo pairwise relationships into a weight matrix $W^{u,t}$;

4: Solve

$$U^t = \arg \min_U L(X_l, U, W^l) + \eta R(X_u, U, W^{u,t})$$
to obtain the new projection
matrix($\eta = (1 - ratio)^{(T-t)}$), ratio is the
utilization of labeled data;

Until: $t > T$ or the stop condition is met.

Output: The projection matrix U .

In accordance with the above grouping method, we conducted randomly ten trails, and finally took the average of ten times as the re-id rate. In person re-id, the cumulative matching characteristic curve is commonly used to describe hit probability in academic. When testing the performance of the centralized test algorithm model, the test sample takes a distance value in turn with N gallery samples, then sort through the order from small to large, and draw the CMC curve from the value of the inter-class sample in the front Top-k. In order to facilitate comparison, the CMC-rank curve is used as the basis of discrimination.

Table 1. Experimental environment

Experimental Parameters	Illustration
CPU	Intel Core CPU 2.20GHz
Memory	4G
Operating System	Windows 7 64 bits
Experimental Platform	MATLAB 8.6.0.267246 (R2015b)

Four images as following are experimental results of four metric learning algorithms on three databases. We should note that we set ratio=1/3 on SSL in all experiments. The ratio equal to 1/3 means that we only use 1/3 of labeled data.

In experiments, we use blue line, red dotted line and black point line represent results on VIPeR, CUHK01 and CUHK03 respectively. It can be clearly seen from Figure 4, no matter how algorithms we applied, no matter we use supervised or semi-supervised learning, the re-identification rates range from large to small are black, red and blue corresponding to CUHK03, CUHK01 and VIPeR. But if we focus on the first-shooting rate (as shown in Figure 5), we will find that the starting point of the red line is always above the black line, and the blue line is always at the bottom. Compared with the characteristics of CUHK: a) image size in CUHK01 are consistent, while CUHK03 are contrary; b) the number of cameras and pedestrians in CUHK03 are far beyond in CUHK01. So when rank 1, CUHK01 is more advantageous that it has faster matching speed and higher recognition rate. While along with the rank increasing, the matching is more easily successful in testing period in CUHK03 because of its large average images per person. So the performance get better than CUHK01 in the middle and posterior segment of the curve. As for VIPeR, its images have lower solution and fuzzy features, so the extraction has more difficulties. The first-shooting rate is only about 40% in VIPeR far below the 60% in CUHK dataset.

The above three charts in Figure 6 are the results of four comparison experiments on three databases. As a whole, the red line is always at the top. Therefore, the MLAPG metric represented by the red line has a more stable re-identification performance on rank 1- rank 100. We analysis the results of three database experiments as following:

The first-shooting performance of XQDA, MLPAG, DNS and SSL in VIPeR increase in turn, up to the extent of 2.63%. But from rank 10 to rank 60, we can see that SSL performance is worse than other three methods. Compared with image feature in VIPeR, because MLAPG applied asymmetric samples weighting strategy based on XQDA, and used APG to accelerate obtain the optimal solution in distance function, so the whole performance on MLAPG is prior to XQDA. While DNS takes SSS into account especially for small database like VIPeR, it is more advantageous.

Experiments in CUHK have some difference with VIPeR, for example, SSL has the worst performance in first-shooting while higher than XQDA after rank 30. Compared with image features in CUHK, the number of person, camera and average images per person are far beyond the VIPeR, so only using part of labeled data like SSL is inferior to supervised methods.

We also compare the performance of these four methods on three databases. From the results shown in Table 2, 3, 4, we observe that we can obtain satisfactory performance by only labeling a small proportion of training data is using an effective semi-supervised learning strategy, especially on CUHK01 database. It is important to note that all experiments of SSL are under ratio = 1/3 circumstance. We can see from Table 2 that SSL has best rank 1 performance with 0.35% higher than DNS. While DNS has better performance than other three algorithms in rank 5 and rank 10 with 71.46% and 82.95% respectively. In Table 3, we observe that SSL has great promising on CUHK01 than MLAPG and DNS both in all ranks with 68.46%, 87.33%, 92.53% and 96.13% respectively.

Figure 4. Results of XQDA, MLAPG, DNS, SSL on three datasets

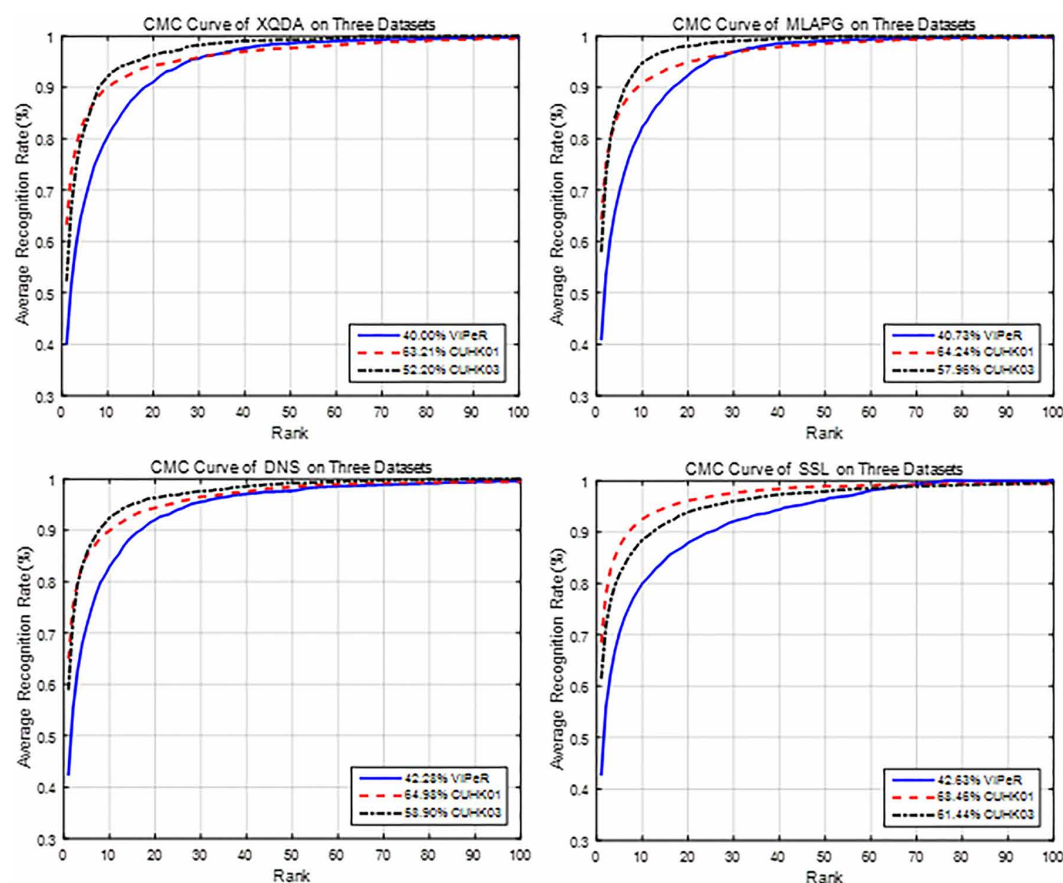


Figure 5. Rank 1 rate of four methods on three datasets

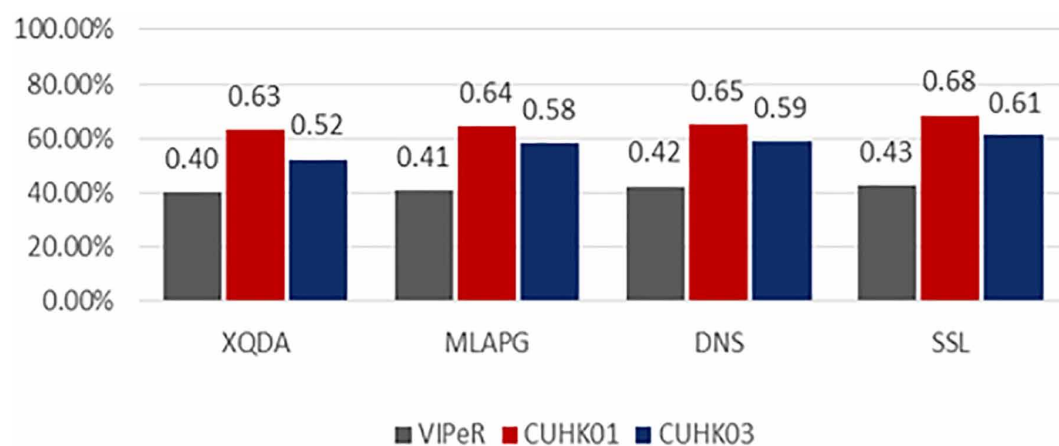
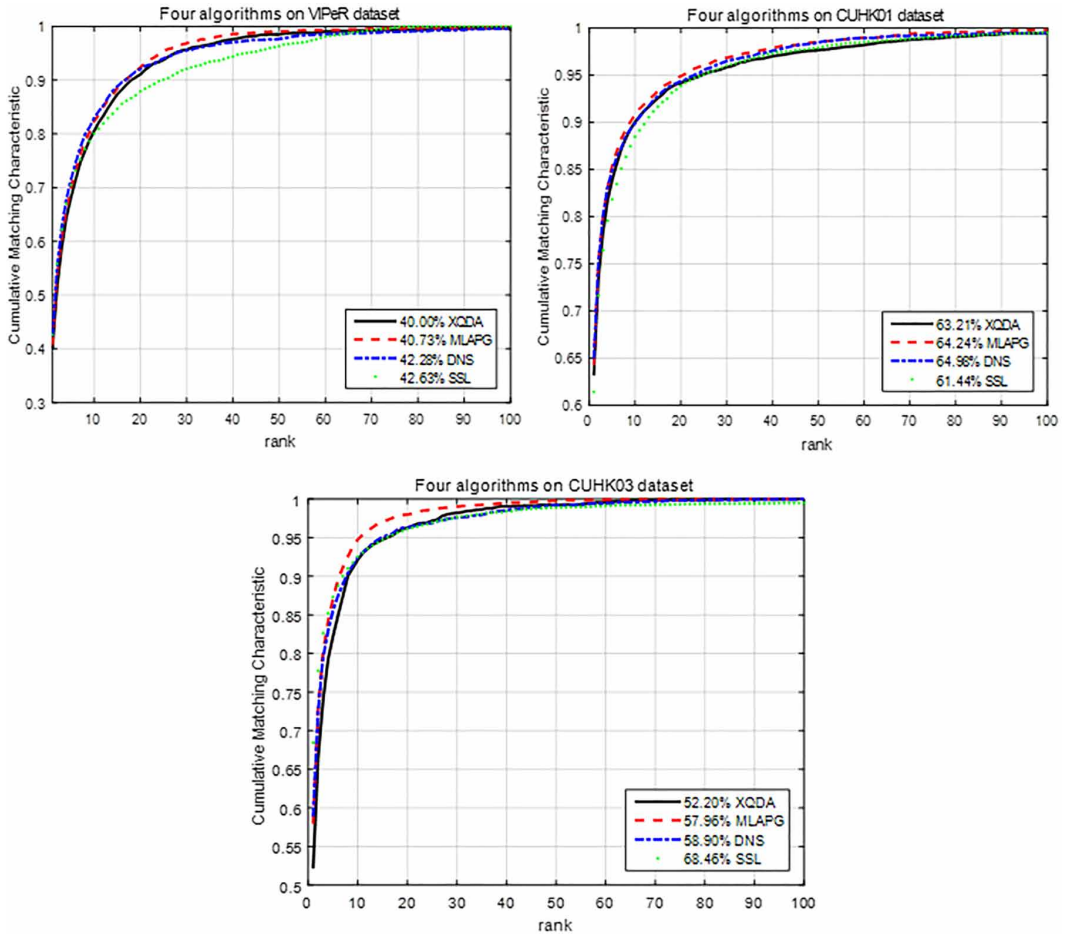


Figure 6. Performance comparison on three datasets



While on CUHK03, SSL only goes beyond on rank 1 with 61.44%, MLAPG has better matching rates on rank 5, rank 10 and rank 20.

In the final, we compare the performance of our approach with SSL on VIPeR, CUHK01 and CUHK03 in Table 5. From the results shown in Table 5, we observe that our approach has a slight promotion compared with SSL on VIPeR, CUHK01 and CUHK03 databases with 0.45%, 0.44% and 0.34% respectively. So, we can confirm that the change of η can improve the matching rate of SSL and get better re-id performance.

6. CONCLUSION

This paper aims at the problem of person re-identification, choosing LOMO feature representation proposed in 2015 CVPR by Liao to match four metric learning algorithms. We can see that MLAPG is prior to XQDA only for large databases. CUHK03 has 13164 person images so it is suitable to apply MLAPG with asymmetric weighting strategy. But take computation into account, MLAPG cost more storage when simulating, so it is practical to apply XQDA in small databases such as VIPeR and CUHK01. DNS is prior to MLAPG only for small data-bases. On the other hand, SSL as only one semi-supervised method, the performances are not very stable. Considering the parameter

Table 2. Performance comparison on the VIPeR database

Methods	Rank=1	Rank=5	Rank=10	Rank=20
XQDA	40.00%	68.13%	80.51%	91.08%
MLAPG	40.73%	69.94%	82.34%	92.37%
DNS	42.28%	71.46%	82.94%	92.06%
SSL	42.63%	70.51%	80.04%	92.92%

Table 3. Performance comparison on the CUHK01 database

Methods	Rank=1	Rank=5	Rank=10	Rank=20
XQDA	63.21%	83.89%	90.04%	94.16%
MLAPG	62.24%	85.41%	90.84%	94.92%
DNS	64.98%	84.96%	89.93%	94.36%
SSL	68.46%	87.33%	92.53%	96.13%

Table 4. Performance comparison on the CUHK03 database

Methods	Rank=1	Rank=5	Rank=10	Rank=20
XQDA	52.20%	82.23%	92.14%	96.25%
MLAPG	57.96%	87.09%	94.74%	98.00%
DNS	58.90%	85.60%	92.45%	96.30%
SSL	61.44%	81.84%	88.48%	93.85%

Table 5. Performance comparison of SSL and ours on three databases (ratio=1/3)

	Methods	Rank=1	Rank=5	Rank=10	Rank=20
VIPeR	SSL	42.63%	70.51%	80.04%	87.92%
	Ours	43.08%	71.18%	80.78%	88.72%
CUHK01	SSL	68.45%	87.32%	92.53%	96.10%
	Ours	68.89%	87.86%	93.09%	96.71%
CUHK03	SSL	61.44%	81.84%	88.48%	93.80%
	Ours	61.78%	82.27%	88.94%	94.33%

between labeled term and unlabeled term in SSL is always constant during iteration, so we propose an improved method which continuously changes the parameter based on the current ratio. And the performance of this modified algorithm is better than SSL original performance. It indicates that semi-supervised metric learning has same matching rate and even better ones. When experiment performance on a small dataset, it has best rate with ratio=1/3. But with the number of image data is increasing, semi-supervised effect is inferior to supervised algorithms. Therefore, if we only focus on whole performance and not consider first-shooting, the semi-supervised should be spread because it has advantages in practical significance and low ratio of labeled data.

In the future, we will strengthen the study of semi-supervised or unsupervised learning, thereby reducing the workload of manual labeling is essential to re-id development. At the same time, we should make full use of the superiority of neural network to further optimize the person re-id algorithm. Speed up the transition from academia to industrial applications.

REFERENCES

- An, S. (2015). Person re-identification by robust canonical correlation analysis. *IEEE Signal Processing Letters*, 22(8), 1103–1107. doi:10.1109/LSP.2015.2390222
- Cai, D., He, X., & Han, J. W. (2007). Semi-supervised discriminant analysis. *Paper presented at the 2007 IEEE International Conference on Computer Vision*. doi:10.1109/ICCV.2007.4408856
- Chapelle, S., & Zien. (2009). Semi-supervised learning. *IEEE Transactions on Neural Networks*, 20(3), 542–542.
- Guo, Y. F., Wu, L. D., Lu, H., Feng, Z., & Xue, X. Y. (2006, November). Null Foley-Sammon transform. In *Pattern recognition*, 39(11), 2248–2251. doi:10.1016
- Hoi, S. C. H., & Lyu, M. R. (2008). A Multimodal and Multilevel Ranking Scheme for Large-Scale Video Retrieval. *IEEE Transactions on Multimedia*, 10(4), 607–619. doi:10.1109/TMM.2008.921735
- Jobson, D. J., Rahman, Z., & Woodell, G. A. (1997). Properties and performance of a center/surround retinex. *Image Processing. IEEE Transactions on*, 6(3), 451–462. PMID:18282940
- Jobson, D. J., Rahman, Z., & Woodell, G. A. (1997). A multi-scale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7), 965–976. PMID:18282987
- Kostinger, M., Hirzer, M., Wohlhart, P., & Bischof, H. (2012, June). Large scale metric learning from equivalence constraints. *Paper presented at the 2012 IEEE Conference on Computer Vision and Pattern Recognition*. doi:10.1109/CVPR.2012.6247939
- Land, E. H., & McCann, J. J. (1971, January). Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1), 1–11. doi:10.1364/JOSA.61.000001 PMID:5541571
- Liao, S. C., Hu, Y., Zhu, X. Y., & Li, S. Z. (2015, June). Person re-identification by local maximal occurrence representation and metric learning. *Paper presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition*. doi:10.1109/CVPR.2015.7298832
- Liao, S. C., & Li, S. Z. (2015, December). Efficient PSD Constrained Asymmetric Metric Learning for Person Re-identification. *Paper presented at the 2015 IEEE International Conference on Computer Vision*. doi:10.1109/ICCV.2015.420
- Liao, S. C., Yi, D., Lei, Z., Qin, R., & Li, S. Z. (2009, June). Heterogeneous face recognition from local structures of normalized appearance. In *International Conference on Biometrics*. doi:10.1007/978-3-642-01793-3_22
- Liao, S. C., Zhao, G. Y., Kellokumpu, V., Pietilainen, M., & Li, S. Z. (2010, June). Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. *Paper presented at the 2010 Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. doi:10.1109/CVPR.2010.5539817
- Loy, C. C., Liu, C., & Gong, S. (2013). Person re-identification by manifold ranking. *Paper presented at the 2013 IEEE International Conference on Image Processing*. doi:10.1109/ICIP.2013.6738736
- Ma, L. Y., Yang, X. K., & Tao, D. C. (2014, August). Person re-identification over camera networks using multi-task distance metric learning. *IEEE Transactions on Image Processing*, 23(8), 3356–3670. doi:10.1109/TIP.2014.2331755 PMID:24956368
- Moghaddam, B., Jebara T. & Pentland A. (2002). Bayesian face recognition. *Pattern Recognition*, 33(11), 1771–1782.
- Nesterov, I. E. (2004). *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers. doi:10.1007/978-1-4419-8853-9
- Pedagadi, J., S., & Boghossian. (2013, June). Local fisher discriminant analysis for pedestrian re-identification. *Paper presented at the 2013 IEEE Conference on Computer Vision and Pattern Recognition*. doi:10.1109/CVPR.2013.426
- Prosser, B., Zheng, W. S., Gong S. G. & X, T. (2010). Person re-identification by support vector ranking. In *BMVC*. doi:10.5244/C.24.21

- Xing, E., Ng, A. Y., Jordan, M., & Russell, S. (2002). Distance metric learning with application to clustering with side-information. In *Proceedings of Neural Information Processing Systems*.
- Yang, X., Wang, M., Hong, R. H., Tian, Q., & Rui, Y. (2017, April). Enhancing Person Re-identification in a Self-trained Subspace. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*. arXiv:1704.06020
- Yang, Y., & Yang, J. M., Y, J. J., Liao, S. C., Yi, Dong. & Li, S. Z. (2014). *Salient color names for person re-identification. Paper presented at the 2014 Proceedings of the European Conference on Computer Vision*.
- Zhang, L., Xiang, T., & Gong, S. G. (2016). Learning a Discriminative Null Space for Person Re-identification. *Paper presented at the 2016 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. doi:10.1109/CVPR.2016.143
- Zheng, S., & Xiang. (2013). Re-identification by relative distance comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3), 653–668.
- Zheng, S. W., Gong, S. G., & Xiang, T. (2011). Person re-identification by probabilistic relative distance comparison. In *Computer Vision and Pattern Recognition*.
- Zhu, X. J. & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, 3(1), 1–130.

Yu Lu received the B.S. degree in communication engineering from Shandong Agricultural University, Taiwan, China, in 2018. She will work toward the M.S. degree in information and communication engineering at Shandong University, Qingdao, China. Her current research interests include machine learning and multimedia data analysis, such as person re-identification under video surveillance.

Lingchen Gu received the B.S degree in electronic information science and technology from Shandong University, Weihai, China, in 2016. She is currently working toward the Ph.D. degree in information and communication engineering at Shandong University, Jinan, China. Her current research interests include machine learning and multimedia data analysis, such as large-scale multimedia indexing and retrieval.

Ju Liu received the B.S. and M.S. degrees in electronic engineering from Shandong University, Jinan, China, in 1986 and 1989, respectively, and the Ph.D. degree in signal processing from Southeast University, Nanjing, China, in 2000. Since July 1989, he has been with the Department of Electronic Engineering, Shandong University, where he is currently a Full Professor. From July 2002 to December 2003, he was a Visiting Professor with the Department of Signal Theory and Communication, Polytechnic University of Catalonia and the Telecommunications Technological Center of Catalonia, Barcelona, Spain. His current research interests include spacetime processing in wireless communications, blind signal separation, and multimedia communications. Dr. Liu is a member of the Editorial Committee of the *Journal of Swarm Intelligence Research* and the *Journal of Electronics (China)*.

Peng Lan received the B.S. and Ph.D. degrees from the College of Information Science and Engineering, Shandong University, Jinan, China, in 2004 and 2009, respectively. Since 2009, he has been an Associate Professor with Shandong Agricultural University, Taiwan, China. Since 2015, he has been an Advanced Visiting Scholar with the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. His research interests include compressive sensing, optimization methods, massive MIMO, and cooperative communications.