Semi-supervised LDA and Multi-distance Metric Learning for Person Re-identification

Bin Li*, Haoyang Ding† and Mengjing Zhou†

1 School of Mechanical Engineering, Wuhan Polytechnic University, Wuhan, Hubei, China
*email: lb420@whpu.edu.cn

Abstract. The problem of person re-identification has attracted a lot of attention in the field of machine vision. In practice, the non-overlapping sample images change drastically and the sample size is small, which makes the metric model overfitting phenomenon. In this paper, based on the k-NN and the sample normality property, we propose a resampling linear discriminant analysis (LDA) algorithm to suppress the local constraints caused by small samples, then train it to obtain the person re-identification metric learning model. A semi-supervised LDA algorithm with semi-supervised characteristics is developed by optimizing the inter-class scatter for weighting. A joint distance metric-based approach is also proposed to learn both the Mahalanobis distance and Euclidean distance. The improved algorithm is tested on the VIPeR and CUHK01 datasets, and the results indicate that, despite the change in the total number of training samples, the algorithm in this paper shows high recognition accuracy.

1. Introduction

In recent years, person re-identification technology has been a hot topic in the research field, taking advantage of the large area of surveillance cameras and the wave of development of computer vision technology[1-5]. The difficulty of person re-identification task lies in the need of authenticated identification of pedestrians in different scenes at different times in a surveillance network with non-overlapping coverage. Compared with traditional image recognition, factors such as lighting, shooting perspective, person posture, and background will person re-identification captured person image appearance features[2,3], and the external features of the same person in the same time period will change dramatically under different external environmental conditions and different perspectives, as shown in Figure 1.

![Person re-identification sample images](image)

**Fig. 1** Person re-identification sample images

Feature extraction, metric model training, and similarity metric are three necessary steps in person re-identification, aiming to continuously improve the discriminative ability and robustness of image feature representation models. In the subsequent research work, the establishment of the sample
distance metric model has received a lot of attention, and researchers have conducted in-depth research on the metric model-based person re-identification algorithm.

Numerous algorithms have been developed based on metric learning, such as RDC[6], ITML[7], KISSME[8], XQDA[9], and NFST[10]. The drastic changes in pedestrian appearance in different camera fields of view, the existing dataset can only cover a few appearance variations, therefore, the recognition accuracy of the person re-identification method based on metric learning has been considerable improved compared to existing studies despite. In spite of great development in person re-identification, the overfitting phenomenon caused by the small amount of data is still difficult to avoid, which usually leads to undesired results.

The completion of the metric learning model in this paper builds on the existing results, and its essence is to carry out research on metric learning. The line person re-identification algorithm mainly improves the generalization ability to solve the overfitting caused by small samples. The contributions of this paper are as follows:

- A new metric projection space learning algorithm is proposed based on the LDA algorithm, and a semi-supervised method is used to improve the learning of the projection matrix to make it more generalizable.
- Combining Mahalanobis and Euclidean distance, a multiple distances metric model for pedestrian re-identification is proposed. A new feature representation method based on dictionary learning is proposed, which is studied together with distance measurement, and the appearance feature model is represented by dictionary.
- The reliability and validity of the algorithm are fully verified by testing in the public databases VIPeR and CUHK01, which provide reference for future related research.

2. Improved joint person re-identification algorithm
The person re-identification metric model is built by a linear discriminant analysis algorithm, while the Fisher discriminant criterion function is reduced to a functional equation with a solution (where the overall scatter is minimal). However, the above optimization model is underdetermined and directly leads to overfitting of the model training results. To improve the generalization ability of the final metric model, this paper introduces a semi-supervised metric learning method as an improvement. According to the literature [11], the equivalent form of the problem can be obtained as follows.

\[
\max J(w) = w^T S_n w \quad s.t. \quad w^T S_p w = 1 \tag{1}
\]

If the existing present set \(\{y_j = x_j^p - x_j^g \mid y_j \in \mathbb{R}^n\}\), \(\{z_j = s_j^p - s_j^g \mid z_j \in \mathbb{R}^n\}\), are the given training set, test set. From equation (1), the following objective mathematical relationship equation is established.

\[
\max J(w) = \frac{w^T S_n w}{w^T S_p w} + \frac{w^T S'_p w}{w^T S'_n w} \tag{2}
\]

where \(S_n, S_p\) represent the dispersion of positive and negative samples, respectively. From Equation (1), it is obtained that the overall dispersion of different samples of the data is bounded by the improved metric learning model to strengthen the generalization ability of the model.

However, since the label information embedded in the test data is not known, in this paper, the semi-supervised learning method is used to perform stepwise optimization calculations in order to solve \(S'_n, S'_p\). The method is shown as follows.

1. Based on the training samples, the Lagrange multiplier method can be used to solve Eq. (1) to obtain the initial metric model and then derive the derivatives on both sides and make the derivatives zero, i.e., the optimization problem in Eq. (1) is transformed into the characteristic root problem.

\[
\left|\lambda I - S^{-1}_p S_n w\right| = 0 \tag{3}
\]
To solve the eigenroot problem in Equation (3), the eigenvector can be solved as \((w_1, w_2, w_3, L, w_n)\). The eigenvalues of \(W = (w_1, w_2, L, w_r)\) in the order of their magnitude are orthogonal to each other. \(w = (w_1, w_2, L, w_r)\) is the projection matrix composed of its first \(r\) terms, i.e., the optimization problem of Equation (1) is solved.

(2) The positive and negative sample scatter \(S'_p, S'_n\) is computed for unlabeled data with the following equations.

\[
S'_p = \sum k_{ij} (z_j (z_j))^T \\
S'_n = \sum (1 - k_{ij}) (z_j (z_j))^T
\]

(4)

where \(\{z_{ij}\}\) denotes the set of test data, \(k_{ij}\) denotes the test samples \(S'^p_j\) and \(S'^n_i\) in the metric subspace with or without k-nearest neighbors.

(3) Self-correction of the metric model. Based on the above calculations, an approximate prediction can be obtained for the unlabeled positive and negative sample scatter. Equation (2) is solved using the Lagrange multiplier method and is transformed into the problem of solving the eigenvalue.

\[
\lambda I - (S_n + S'_n)^{-1} (S_p + S'_p) W = 0
\]

(5)

The projection matrix \(W' = (w'_1, w'_2, L, w'_r)\) of the algorithm metric space in this paper can be obtained by calculating the characteristic roots in Equation (5). Then, using the learned metric model, the similarity metric \(D_{w'}\) of the test samples can be solved and the test samples can be ranked based on the above results.

3. Person re-identification based on multi-distance metric learning

3.1. A linear discriminant analysis based person re-identification method

A metric model based on multiple distances is established by first extracting the appearance features of pedestrian images using the LOMO model, and forming sample pairs by randomly selecting each sample in the sample set according to the label information. As described in the literature [11], the metric model can be described as Equation (6).

\[
d (x^p, x^p) = (x^p - x^p)^T W^T M W (x^p - x^p)
\]

(6)

where \(W\) denotes the projection matrix of the mapping subspace, and \(M\) is the metric matrix of the Mahalanobis distance. The metric-based learning approach aims to bring the positive samples in the learning metric subspace \((W, M)\) closer to each other than the negative sample pairs.

In addition, we propose a new dictionary representation feature in this model to improve the Euclidean distance. The new features are learned through a lexical representation. The multi-distance model of the described method, as shown in Equation (7).

\[
d (x^p, x^p) = (x^p - x^p)^T W^T M W (x^p - x^p) + \|u^p - u^p\|
\]

(7)
where $\textbf{u}_i^p$ and $\textbf{u}_j^g$ denote features corresponding to the dictionary representations of $\textbf{x}_i^p\textbf{W}$ and $\textbf{x}_j^g\textbf{W}$, respectively. We use the XQDA[9] method to learn feature projections. Then, the fisher discriminant criterion is expressed as shown in Equation (8).

$$
\max_w J(w) = \frac{w^T S_b w}{w^T S_w w}
$$

where $S_b$ is the interclass scatter and $S_w$ is the intraclass scatter, as described in the literature [11], the above equation is simplified based on the Lagrange multiplier method to solve the equational generalized eigenvalue problem, and the KISSME[8] method is used to learn the Mahalanobis distance to obtain the distance similarity metric. The Mahalanobis distance can be got by Equation (9).

$$
M = \left( (\textbf{W}\Sigma_p \textbf{W}^T)^{-1} - (\textbf{W}\Sigma_n \textbf{W}^T)^{-1} \right)
$$

where $\Sigma_p$ is the covariance of positive sample pairs and $\Sigma_n$ is the covariance of negative sample pairs. Therefore, $\textbf{W}\Sigma_p \textbf{W}^T$ and $\textbf{W}\Sigma_n \textbf{W}^T$ are the corresponding covariances in the metric subspace.

3.2. Combining dictionary representation learning with a unified dictionary

To describe the intrinsic relationship between cross-camera images, it is assumed that the appearance features of all people can be expressed in the same dictionary, and the features of people in cameras have different coefficients. It can be conjectured that for the same person, there ought to be a mapping function to convert the coefficients in camera A to the coefficients in camera B.

For a given training set $\textbf{X}^p, \textbf{X}^g, \textbf{X}^p, \textbf{X}^g \in \mathbb{R}^{d \times n}$, where $\textbf{X}^p$ is the image set of camera A, $\textbf{X}^g$ is the image set of camera B, $\textbf{x}_i^p = \{x_i^p, x_i^p, \ldots, x_i^p\}$, $\textbf{x}_i^g = \{x_i^g, x_i^g, \ldots, x_i^g\}$, $d$ is the dimensionality of the eigenvector $\textbf{x}_i^p$. Let $\textbf{E} \in \mathbb{R}^{k \times k}$ denote the joint feature dictionary, $\textbf{F} \in \mathbb{R}^{k \times k}$ denote the mapping matrix connecting the individuals in the problem set to the image set of camera B, and $k$ denote the dimension of the features represented by the dictionary. That is, the projected features in the metric subspace can be decomposed into Eq. (10), and Eq. (11) is obtained:

$$
\textbf{X}^p\textbf{W} = \textbf{E}\textbf{U}, \textbf{X}^g\textbf{W} = \textbf{E}\textbf{V},
$$

$$
\textbf{V} = \textbf{F}\textbf{U}
$$

where $\textbf{U} \in \mathbb{R}^{k \times N}$, $\textbf{V} \in \mathbb{R}^{k \times N}$ are the dictionary representation features of camera A and B image sets, respectively.

The goal is to minimize the reconstruction error of the dictionary representation and the sample matching error of the intermediate layer feature representation. The objective function of the optimization algorithm is shown in Equation (12).

$$
\min_{\textbf{E}, \textbf{F}} \begin{bmatrix}
\|\textbf{X}^p\textbf{W} - \textbf{E}\textbf{U}\|^2 + \alpha\|\textbf{X}^g\textbf{W} - \textbf{E}\textbf{V}\|^2 \\
+ \lambda\|\textbf{F}\textbf{U} - \textbf{U}\|^2 + \rho\|\textbf{F}\|^2 + \|\textbf{V}\|^2
\end{bmatrix},
$$

subject to $\|\textbf{U}\| \leq 1$, $\|\textbf{F}\| \leq 1$, $\forall i$

where $\textbf{e}_i$ and $\textbf{f}_i$ correspond to the i-th columns of $\textbf{E}$ and $\textbf{F}$, respectively. $\alpha$, $\lambda$ and $\rho$ are the equilibrium parameters of the corresponding terms.
3.3. Model training

The joint dictionary representation learning optimization algorithm for solving Equation (12) is summarized as follows.

1. Initialize $\mathbf{E}, \mathbf{F}$ and $\mathbf{V}$ with random matrices, respectively.
2. Fixing $\mathbf{E}, \mathbf{F}$ and $\mathbf{V}$, then $\mathbf{U}$ can be calculated by solving equation (13).

$$\mathbf{U} = (\mathbf{E}'\mathbf{E} + \lambda\mathbf{F}'\mathbf{F} + \rho\mathbf{I})^{-1} (\mathbf{E}'\mathbf{X} + \lambda\mathbf{F}'\mathbf{V})$$

(13)

3. Similarly, fixing $\mathbf{E}, \mathbf{F}$ and $\mathbf{U}$, calculate $\mathbf{V}$ by solving equation (14).

$$\mathbf{V} = (\alpha\mathbf{E}'\mathbf{E} + \lambda\mathbf{I} + \rho\mathbf{I})^{-1} (\alpha\mathbf{E}'\mathbf{V} + \lambda\mathbf{F}\mathbf{U})$$

(14)

4. To facilitate the solution, we use the Lagrangian pairwise algorithm to learn $\mathbf{E}$. After fixing the other variables, the sparse coding method in [12] is used to obtain $\mathbf{E}$ with the mathematical expression shown in Equation (15).

$$\mathbf{E}' = (\mathbf{UU}' + \alpha\mathbf{VV}' + \lambda\mathbf{I})^{-1} (\mathbf{X}'\mathbf{U}' + \alpha\mathbf{X}'\mathbf{V}')$$

(15)

5. Similarly, the Lagrangian pairwise algorithm can be used to learn $\mathbf{F}$, which has a similar form to $\mathbf{F}$ in Eq. (12) and whose expression is shown in Eq. (16).

$$\mathbf{F}' = (\mathbf{UU}' + \lambda\mathbf{I})^{-1} (\mathbf{VV}')$$

(16)

4. Simulation experiments

4.1. Database and evaluation metrics

In this paper, VIPeR[3] and CUHK person re-identification datasets are selected for simulation experiments to verify the effectiveness of the algorithm.

The CMC cumulative accuracy curve[5,12] is used to evaluate the accuracy of pedestrian re-recognition in this paper, which is derived from the calculation of the ratio of correct results to the total number to be recognized among the top-ranked recognition results.

4.2. Simulation experiments

In this paper, the dataset is randomly divided to generate training samples and test samples to evaluate the recognition accuracy of the algorithm at the same training scale, and the results of 10 independent repetitions are averaged as the final recognition results. The results are shown in Table 1 and Table 2, where $p$ denotes the training sample size.

Table 1 shows the comparison test results of the improved LDA algorithm and the multi-distance metric learning algorithm on the VIPeR database in this paper, which is quantified by recognition accuracy statistics of rank-1, rank-5 and rank-10. From the test results, it can be seen that the recognition accuracy of the two algorithms proposed in this paper are significantly improved compared with the existing state-of-the-art algorithms. On rank-1, the recognition accuracy of the improved LDA algorithm in this paper is improved by 2.88% compared with the best comparison algorithm. The improved LDA algorithm improved the recognition accuracy by 6.29% compared to the pre-improvement. The multi-distance metric learning algorithm improves 4.88% and 3.80% on rank-5 and rank-10 metrics, respectively, compared with the best comparison algorithm. The recognition accuracy of improved LDA algorithm in these two metrics also almost equal to that of the existing state-of-the-art algorithms. Its CMC curve is shown in Figure 2.

| Tab. 1 VIPeR database comparison experiment recognition accuracy statistics table |
|-----------------------|-----------------|-----------------|-----------------|
| methods               | $p=316$         |                 |                 |
|                      | rank=1          | rank=5          | rank=10         |
| KISSME[8]             | 19.94           | 47.89           | 62.77           |
| XQDA[9]               | 40.00           | 68.33           | 81.07           |
Table 2 then gives the comparison results of both algorithms with three advanced algorithms such as XQDA[9], NFST[10] and MLAPG[13] on CUHK01 database. It is obvious both algorithms in this paper have excellent performance and achieve the highest level of recognition accuracy in all three metrics shown. The improved LDA algorithm achieves 70.00% recognition accuracy in rank-1, which exceeds the best comparison algorithm by 5.00%, while the multi-distance metric learning algorithm also achieves optimal or near-optimal recognition accuracy in all metrics. In addition, the recognition accuracy of the improved LDA algorithm proposed in this paper is significantly improved compared with the original algorithm, which shows that the improvements made by the improved LDA algorithm are quite remarkable.

Comparing Tables 1 and 2, it can be seen that even in different datasets, all algorithms exhibit a positive correlation between the training sample size and recognition accuracy, except for the poorly performing LDA algorithm. Accordingly, we can see that the recognition accuracy of the multi-distance metric learning algorithm decreases less when the training sample size is reduced compared to the XQDA and MLAPG algorithms. In other words, the multi-distance metric learning algorithm can achieve higher recognition accuracy with a smaller training sample size than the comparison algorithms.

5. Conclusion
Aiming at the popular person re-identification problem, this paper establishes an LDA algorithm based on normality resampling. A novel generalized weighted person re-identification algorithm is also proposed by combining the generalized model after resampling with the LDA model. Meanwhile, a person re-identification algorithm based on multi-distance metric learning is established by jointly learning the Mahalanobis and Euclidean distance. Compared with algorithms such as KISSME,
XQDA and MLAPG, the effectiveness, accuracy and adaptability of the two algorithms to size of training samples are fully demonstrated. In terms of rank-1 recognition accuracy, the improved LDA algorithm improves 2.88% and 5.00%, respectively, over the optimal results of other selected algorithms in the described database, and compared with the pre-improvement by 6.29% and 26.77%, respectively. In rank-5 and rank-10 metrics, the multi-distance metric learning algorithm improves by 4.88% and 3.80%, respectively, compared to the best comparison algorithm. In particular, the multi-distance metric learning algorithm can achieve higher recognition accuracy with a smaller training sample size. It can be seen that the two algorithms can be used together to form a complementary advantage for the practical application in different scenarios.

Acknowledgments
The project supported by Scientific Research Project of Hubei Province Department of Education (Q20201601) and Research and Innovation Initiatives of WHPU (2020Y11), this work was finished at Wuhan Polytechnic University, Wuhan.

References
[1] Liu X, Tao D, Song M, et al. 2015 Learning to track multiple targets IEEE Trans on Neural Networks & Learning Systems 26(5) p 1060
[2] Yang Y, Yang J, Yan J, et al. 2014 Salient color names for person re-identification European Conference on Computer Vision pp 536–551
[3] Gray D and Tao H 2008 Viewpoint invariant pedestrian recognition with an ensemble of localized features European Conference on Computer Vision pp 262-275
[4] Zhang Z, Chen Y and Saligrama V 2014 A Novel Visual Word Co-occurrence Model for Person Re-identification Workshop at the European Conference on Computer Vision Springer International Publishing pp 122-133
[5] Hirzer M, Roth P M, Kostinger M, et al. 2012 Relaxed pairwise learned metric for person re-identification European Conference on Computer Vision pp 780-793
[6] Zheng W, Gong S, Xiang T 2013 Reidentification by relative distance comparison IEEE Trans on Pattern Analysis and Machine Intelligence 35(3) pp 653–668
[7] Davis J V, Kulis B, Jain P, et al. 2007 Information-theoretic metric learning International Conference on Machine Learning pp 209–216
[8] Köstinger M, Hirzer M, Wohlhart P, et al. 2012 Large scale metric learning from equivalence constraints IEEE Conference on Computer Vision and Pattern Recognition pp 2288–2295
[9] Zhang L, Xiang T and Gong S 2016 Learning a Discriminative Null Space for Person Re-identification Computer Vision and Pattern Recognition pp 1239-1248
[10] Prosser B, Gong S, and Xiang T 2008 Multi-camera matching under illumination change over time Workshop on Multi-Camera and Multi-Modal Sensor Fusion Algorithms and Applications
[11] Song L, Li B, Zhao J and Liu G 2020 Normality Resampling of Improved Metric Learning Method for Person Re-Identification Computer Engineering and Applications 56(08) pp 158-165
[12] H. Lee, A. Battle, R. Raina and A.Y. Ng 2007 Efficient sparse coding algorithms Advances in neural information processing systems pp 801-808
[13] Ahmed E, Jones M and Marks T K 2015 An improved deep learning architecture for person re-identification Computer Vision and Pattern Recognition pp 3908-3916