Regularized Bayesian Metric Learning for Person Re-Identification

Venice Erin Liong[1], Jiwen Lu[1], and Yongxin Ge[2,3⋆]

[1]Advanced Digital Sciences Center, Singapore
[2]Key Laboratory of Dependable Service Computing in Cyber Physical Society Ministry of Education, Chongqing, 400044, China
[3]School of Software Engineering, Chongqing University, Chongqing, 400044, China

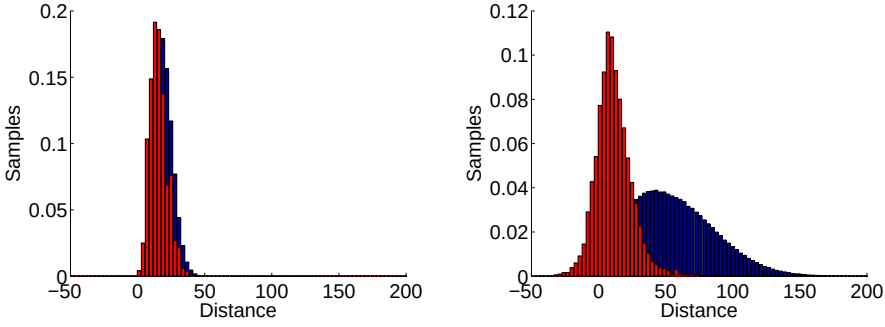
Abstract. Person re-identification across disjoint cameras has attracted increasing interest in computer vision due to its wide potential applications in visual surveillance. In this paper, we propose a new regularized Bayesian metric learning (RBML) method for person re-identification. While numerous metric learning methods have been proposed for person re-identification in recent years, most of them suffer from the small sample size (SSS) problem because there are not enough training samples in most practical person re-identification systems, so that the within-class and between-class variations can be well estimated to learn the distance metric. To address this, we propose a RBML method to model and regulate the eigen-spectrums of these two covariance matrices in a parametric manner, so that discriminative information can be better exploited. Experimental results on three widely used datasets demonstrate the advantage of our proposed RBML over the state-of-the-art person re-identification methods.

Keywords: Person re-identification, metric learning, regularization.

Person re-identification aims to match and recognize persons who have been observed over different disjoint cameras, and it has many potential applications such as information retrieval and visual surveillance. Over the past decade, many person re-identification methods have been proposed in the literature, and they can be mainly classified into two categories: feature-based [1,2] and model-based [3–5]. Feature-based approach aims to extract robust and discriminative features to characterize the appearance of human body images [1, 2, 6–10], and model-based approach learns a classifier or ranker to recognize the query body image.

Recent years have witnessed that metric learning has been one of most popular methods in person re-identification [11–22] . Metric learning aims to learn a Mahalanobis distance metric to transform samples from the original space to another feature space by exploiting the between-class and within-class variations. Representative metric learning algorithms include large margin nearest neighbor (LMNN) [23], information theoretic metric learning (ITML) [24], and neighborhood component analysis (NCA) [25]. The objective of metric learning is to learn a discriminative similarity to better measure the similarity of human body images by making use of some prior knowledge. Fig. 1 illustrates an example to show the advantage of metric learning

⋆ Corresponding author.



(a) Before metric learning. (b) After metric learning.

Fig. 1. Similarity distribution over 3160 positive and 3160 randomly sampled negative pairs from the VIPeR dataset. (a) The original distribution in the Euclidean space. (b) The new distribution in the learned metric space. The histograms with the red color and the blue color represent the matched (same person) and mismatched (different person) image pairs, respectively. Before metric learning, the similarity of matched and mismatched pairs overlaps heavily, which is more challenging for re-identification. However, these two similarity distribution histograms become more separate in the learned distance metric, so that more discriminative information is exploited.

for person re-identification. While metric learning methods have shown promising performance in person re-identification, most of them suffer from the small sample size (SSS) problem because there are usually not enough training samples to estimate the within-class and between-class variations.

In this paper, we propose a new regularized Bayesian metric learning (RBML) method for person re-identification. Our method models the difference of each positive pair and that of each negative pair as a Gaussian distribution model under the Bayesian framework, and optimizes their probability ratio to learn a discriminative distance metric. To better estimate the intra-class and inter-class variations when there are small number of training samples, we decompose the eigen-spectrum of each covariance matrix into two subspaces and regularize them in a parametric manner. By doing so, the learned distance metric can extract more relevant information. Our method is evaluated on three widely used person re-identification datasets and experimental results demonstrate the efficacy of the proposed method.

1 Related Work

Person Re-Identification: Existing person re-identification methods can be categorized into two classes: feature-based and model-based. Feature-based methods extract visual signatures to differentiate different persons. For example, Farenzena *et al.* [1] developed a symmetry driven accumulation of local feature (SDALF) for body appearance modeling. Jungling *et al.* [26] designed a codebook learning approach to cluster local features for implicit body shape matching. Ma *et al.* [27] developed a BiCov descriptor which combines biologically-inspired features and covariance features for human body

representation. Ma *et al.* [28] used local descriptors from pixel intensities and encoded them into higher-order Fisher vectors for feature representation. Gray *et al.* [9] selected a subset of color and texture features for person body representation. Schwartz *et al.* [29] performed partial least squares (PLS) to adaptively weight different local features for body representation. Model-based methods learn a classification or ranking model to recognize the query body image from the gallery set. Representative methods include support vector machine [30], transfer learning [31], manifold ranking [32] and metric learning [5, 25, 33, 34]. In this work, we contribute to the second category and propose a new metric learning approach for person re-identification.

Metric Learning: Metric learning aims to learn a Mahalanobis distance metric to transform samples to another feature space by exploiting the between-class and within-class variations. In recent years, metric learning has been widely used in person re-identification and has also achieved the state-of-the-art performance. For example, Hirzer *et al.* [35] proposed a relaxed pairwise metric leaning approach. Kostinger *et al.* [5] learned a Mahalanobis distance metric based on the statistical inference over the ratio of similar and similar pairs in the training set. Zheng *et al.* [36] proposed a relative distance comparison (PRDC) metric learning to maximize the probability of a positive pair having a smaller distance than a negative pair. However, there are still two shortcomings among these methods: 1) most of them rely on a large number of training samples and may not perform well when the training set is small; 2) most of them perform an iterative optimization procedure for training, which increases the computational complexity. To address this, we propose a simple yet effective metric learning method for person re-identification, where the difference of each pair of positive and negative pairs are represented as a Bayesian model, respectively, and the ratio of positive pairs over negative pairs is minimized. Since the number of training samples is usually small, we regularize the covariance matrices in a parametric manner to make them well-estimated and stable. By doing so, relevant information can be extracted from the learned distance metric.

2 Regularized Bayesian Metric Learning

Let $x \in R^d$ be the feature representation of a body image, $\mathcal{S}$ and $\mathcal{D}$ two sets of similar and dissimilar pairs, where $(x_i, x_j) \in \mathcal{S}$ is a feature pair from the same person and $(x_i, x_j) \in \mathcal{D}$ is a pair from different persons. We compute the following ratio:

$$\delta(x_i, x_j) = \log\left(\frac{\Pr[(x_i, x_j) \in \mathcal{D}]}{\Pr[(x_i, x_j) \in \mathcal{S}]}\right) \tag{1}$$

where $\Pr[\cdot]$ is a probability distribution function to measure the likelihood of whether (x_i, x_j) is from the same person. Specifically, $\delta(x_i, x_j)$ is high if they are from different persons, and low from the same person. We model the probability distribution function as a single Gaussian distribution of ($x_{ij} = x_i - x_j$) with zero mean as follows:

$$\Pr[\cdot] = \frac{1}{\sqrt{2\pi \mid \Sigma \mid}} \exp\left(-\frac{1}{2} x_{ij}^T \Sigma^{-1} x_{ij}\right) \tag{2}$$

Then, (1) can be simplified as follows by removing the log operator and the constant term:

$$\begin{aligned}\delta(x_i, x_j) &= x_{ij}^T \Sigma_1^{-1} x_{ij} + \log(|\Sigma_1|) - x_{ij}^T \Sigma_0^{-1} x_{ij} - \log(|\Sigma_0|) \\ &= x_{ij}^T (\Sigma_1^{-1} - \Sigma_0^{-1}) x_{ij} \end{aligned} \tag{3}$$

where

$$\Sigma_1 = \sum_{(x_i, x_j) \in \mathcal{S}} (x_i - x_j)(x_i - x_j)^T \tag{4}$$

$$\Sigma_0 = \sum_{(x_i, x_j) \in \mathcal{D}} (x_i - x_j)(x_i - x_j)^T \tag{5}$$

These two covariance matrices cannot be well estimated if there are not enough training samples. To address this, we regulate each of them. We first obtain the eigen-spectrum of each covariance matrix as follows:

$$\Sigma = \Phi \Lambda \Phi^T \tag{6}$$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_d)$ denotes the eigenvalues of Σ, which is sorted in a descending order, and $\Phi = [\phi_1, \phi_2, \cdots, \phi_d]$ is the corresponding eigenvector.

The eigen-spectrum obtained with small number of training samples is usually biased because the energy of the spectrum is concentrated at a few principal components corresponding to the largest eigenvalues and those corresponding to smaller eigenvalues are unreliable. To better estimate the eigen-spectrum in the whole space, we divide the original eigen-spectrum of Λ into two subspaces, called principal (P) and noise (N) space and regulate each of them separately:

$$\lambda_t \in \begin{cases} P, & t \leq c \\ N, & c+1 \leq t \leq d \end{cases} \tag{7}$$

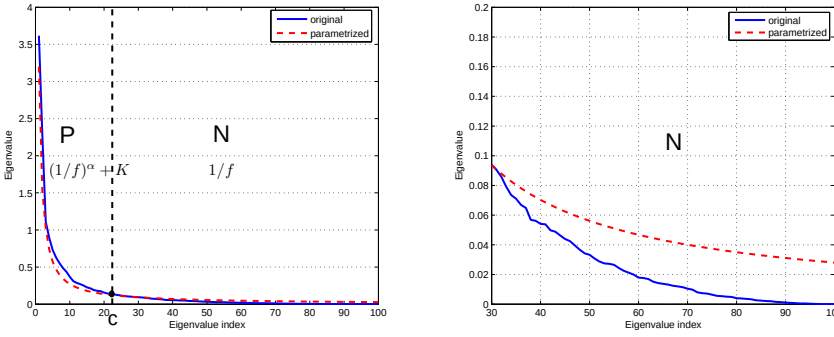
where c is a pre-specified energy percentage parameter which is computed as follows:

$$c = \min\{c | (\sum_{t=1}^{c} \lambda_t / \sum_{t=1}^{d} \lambda_t) \geq \zeta\} \tag{8}$$

To better estimate the eigen-spectrum, we present a new regularization model which reduces the effect of larger eigenvalues in P and enhances the effect of smaller eigenvalues in N as follows:

$$\lambda_t' = \begin{cases} \left(\dfrac{a}{t+b}\right)^{\alpha} + K & \text{for } t \leq c \\ \dfrac{a}{t+b} & \text{for } c+1 < t < d \end{cases} \tag{9}$$

In (9), we regularize the whole eigen-spectrum in a parametric manner by using the $1/f$ function because this function fits well to the decaying nature of the eigen-spectrum [37, 38]. Since the eigenvalues in P are much larger than those in N, we



(a) (b)

Fig. 2. (a) The basic idea of our proposed RBML, where the eigen-spectrum of the covariance matrix is decomposed into two subspaces and regularized with different models. (b) Zooming in of the original and modeled noise spaces.

further shrink and suppress them by using an exponential function using parameter α, where $0 < \alpha < 1$. The parameter K in (9) is used to make the eigen-spectrum continuous in the whole space so that a smooth eigen-spectrum can be obtained in the whole space, where

$$K = \frac{a}{c+b} - \left(\frac{a}{c+b}\right)^{\alpha} \tag{10}$$

Let $\lambda_1 = \frac{a}{1+b}$ and $\lambda_c = \frac{a}{c+b}$. The coefficients a and b are computed as:

$$a = \frac{\lambda_1 \lambda_c (c-1)}{\lambda_1 - \lambda_c} \tag{11}$$

$$b = \frac{c\lambda_c - \lambda_1}{\lambda_1 - \lambda_c} \tag{12}$$

Having obtained the modeled eigen-spectrum, we compute the modeled covariance matrix as follows:

$$\Sigma' = \Phi \Lambda' \Phi^T \tag{13}$$

where $\Lambda' = \text{diag}(\lambda'_1, \lambda'_2, \cdots, \lambda'_d)$ are the eigenvalues which are sorted in a decreasing order. Now, the Mahalanobis distance metric in (3) can be computed as follows:

$$\begin{aligned} \delta(x_{ij}) &= x_{ij}^T (\Sigma_1'^{-1} - \Sigma_0'^{-1}) x_{ij} \\ &= (x_i - x_j)^T \mathbf{M} (x_i - x_j) \end{aligned} \tag{14}$$

where $\mathbf{M} \triangleq \Sigma_1'^{-1} - \Sigma_0'^{-1}$.

Fig. 2 show the estimated and original eigen-spectrum of one covariance matrix, and we see that effect of the eigenvalues in the N space is enhanced in the modeled space. **Algorithm 1** summarizes the proposed RBML method.

Input: Positive and negative training sample pairs and parameters α and ζ.
Output: Mahalanobis distance metric **M**.
Step 1 (Initialization):
1.1. Compute covariance matrices for similar and dissimilar pairs, respectively.
Step 2 (Regularization and Modeling):
For each covariance matrix, repeat
2.1. Perform eigendecomposition using (6).
2.2. Compute for parametric c using (8).
2.3. Estimate the new eigenspectrum using (9).
2.4. Compute the new covariance using (13).
Step 3 (Output):
Output the matrix **M** using (14).

Algorithm 1: RBML

3 Experiments

We evaluate the proposed RBML method on three widely used person re-identification datasets, namely the VIPeR [39], ETHZ [40] and i-LIDS [41] databases. The following describe the details of our experiments and results.

3.1 Settings

For each person image, we extracted a mixture of color and texture histogram features by following the detailed settings in [3, 9]. Each person image is represented by a 2784-dimensional feature vector. Specifically, we divided the person image into six horizontal stripes. For each stripe, the RGB, YCbCr, HSV color features and texture features (Schmid and Gabor) were extracted and represented as a histogram feature. We applied PCA to project the feature representation into a 100-dimensional feature vector to remove the redundancy of the high-dimensional feature space.

Having extracted the feature of each image, we learned a discriminative distance metric by using the proposed RBML method. Finally, the nearest neighbor is used for ranking.

3.2 Evaluation on the VIPeR dataset

The VIPeR dataset [39] consists of 632 persons and each person has two images captured from two different cameras in an outdoor environment. Most image pairs in this dataset have a viewpoint change of 90 degrees, where one is from the front/back view and the other is from the side view. There are large variations of illumination, viewpoint and pose in these captured images. Fig. 3 shows some sample images in the VIPeR dataset.

We followed the experimental settings in [3, 9]. In our experiments, we randomly selected p persons for training and the remaining persons were used for testing. We repeated this selection 10 times and took the average as the final re-identification accuracy. In the training phase, two images from the same person form a positive pair, and

Fig. 3. Sample images of the VIPeR dataset. The images from each column are body images captured from different cameras and belong to the same person.

Table 1. Matching rates (%) of different metric learning methods on the VIPeR dataset.

Method	$p = 316$				$p = 100$			
Rank	1	10	25	50	1	10	25	50
RBML	**27.22**	**72.78**	**89.08**	**95.89**	**13.06**	**47.46**	**64.94**	**79.70**
KISSME [5]	22.15	66.77	82.91	92.41	1.41	7.14	12.78	21.24
RS-KISSME [4]	22.31	67.41	86.23	94.62	10.71	40.98	59.30	76.79
Mahalanobis	13.29	48.58	71.04	84.02	8.93	34.49	52.82	67.86
Identity	4.43	16.93	27.06	42.56	1.41	7.42	12.69	21.43
ITML [24]	0.95	8.23	18.20	32.28	3.67	11.94	20.39	30.08
LMNN [23]	18.04	59.34	78.16	89.72	8.08	34.40	51.60	66.82

two images from different persons form a negative pair. In the testing phase, we used the single-shot testing where one image per person was randomly selected to form the gallery set and the remaining images were used to form the probe set.

We used the cumulative matching curve (CMC) as the evaluation metric in our experiments, where a match is found at the top-n ranks. The parameters ζ and α were empirically set to 90% and 0.9, respectively.

Comparison with Existing Metric Learning Algorithms: Table 1 tabulates the matching rates of different metric learning methods, which includes LMNN, ITML, KISSME and RS-KISSME, as well as the widely used Mahalanobis and Identity (Euclidean) distance metrics on the VIPeR dataset. We see that our proposed RBML method consistently outperforms all metric learning methods with as high as 5% and 2% Rank-1 accuracy at $p = 316$ and $p = 100$, respectively. Fig. 4 shows the matching examples of our RBML method on the VIPeR dataset.

Comparison with State-of-the-Art Person Re-identification Methods: We also compared our RBML with the state-of-the-art person re-identification methods on the VIPeR dataset. Tables 2 and 3 tabulate the matching accuracies on this dataset when p is set to 316 and 100, respectively. We see that RBML outperforms all the state-of-the-art person re-identification across several ranks. Particularly, the rank-1 matching rate of

Fig. 4. Examples of person re-identification on the VIPeR dataset using our proposed RBML. The first column indicates the probe images, the middle area shows the top-15 ranked results from the gallery image with a highlighted red box for the correct match with our method, and the last column shows the true match.

Table 2. Matching rates (%) of our method and the state-of-the-art person re-identification methods on the VIPeR dataset when p is set as 316.

Method	r=1	r=5	r=10	r=20
RBML	**27.53**	**58.23**	**71.52**	**84.65**
eSDC [10]	26.74	50.70	62.37	76.36
eLDFV [28]	22.34	47.00	67.04	71.00
eBiCov [27]	20.66	42.00	56.18	68.00
CPS [42]	21.84	44.00	57.21	71.00
SDALF [1]	19.87	38.89	49.37	65.73
ELF [9]	12.00	31.00	41.00	58.00
PRDC [3]	15.66	38.42	53.86	70.09
aPRDC [43]	16.14	37.72	50.98	65.95
PCCA [33]	19.27	48.89	64.91	80.28

our method improves the current state-of-the-art methods by 0.8% and 2.1% when p is set as 316 an 100, respectively.

Parameter Analysis: We also investigated the individual contributions of our RBML model. We define three variations of our method to study their different importance: 1) RBML1: regularizing the noise space only; 2) RBML2: regularizing without the suppression in the principal space; and 3) RBML3: regularizing the noise and principal space without continuity. RBML1 was implemented by regularizing the noise space with the $1/f$ function only and the principal space is not regularized. RBML2 ignored the suppression in the principal space, which is equivalent to setting the parameter α as 1 in RBML. RBML3 performed the same modeling method as RBML for both the principal and noise spaces but discarded the continuity parameter K in the principal space. Fig. 5 shows the matching rates of different types of regularization model in our RBML on the VIPeR dataset. It can be seen that removing a portion of our model degrades the

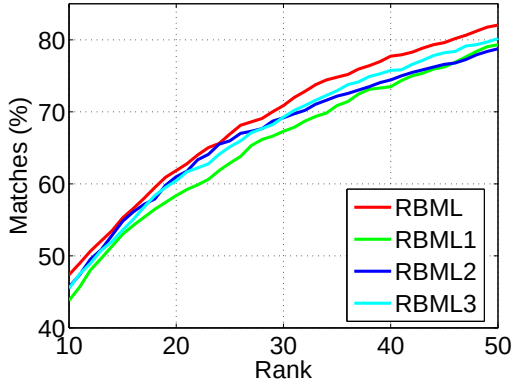


Fig. 5. CMC curves for different types of regularization methods in our model at $p = 100$ on the VIPeR dataset.

Table 3. Matching rates (%) of our method and the state-of-the-art person re-identification methods on the VIPeR dataset when p is set as 100.

Method	r=1	r=5	r=10	r=20
RBML	**13.16**	**34.12**	**46.71**	**60.62**
Hirzer's [35]	11.00	-	38.00	52.00
PRDC [3]	9.12	24.19	34.40	48.55
PCCA [33]	9.27	24.89	37.43	52.89
MCC [25]	5.00	16.32	25.92	39.64

whole re-identification performance. Moreover, the regularization in the noise space is the most important in our method and the suppression in the principal space also contributes to the overall performance. Lastly, the continuity of the eigen-spectrum is also useful to improve the re-identification result.

We also investigated the performance of our method versus varying values of α. Fig. 6 shows the CMC curves versus varying values of α. We see that the performance of our method degrades heavily when α is smaller than 0.9.

Computational Time: We investigated the computational time of different methods on the VIPeR dataset. Our hardware configuration comprises a 2.4-GHz CPU and a 8GB RAM. Table 4 shows the average training time of different methods on the VIPeR dataset when the number of p is set as 316. We see that our RBML method is more efficient than other popular metric learning methods such as ITML and LMNN, and as efficient as KISSME. That is because our regularization operation doesn't require any complex optimization and iteration.

Comparison with Other Regularization Methods: We also compared our method with other regularization models such as smoothing and shrinkage. For smoothing regularization, the noise space of the eigen-spectrum of the covariance matrix is defined by a constant, $\beta = \frac{1}{d-c}\sum_{t=c+1}^{d}\lambda_t$. The shrinkage regularization model [44] regulates the model as: $\Sigma' = (1-\gamma)\Sigma + \gamma\tau\mathbf{I}$, where γ is the shrinkage parameter which ranges

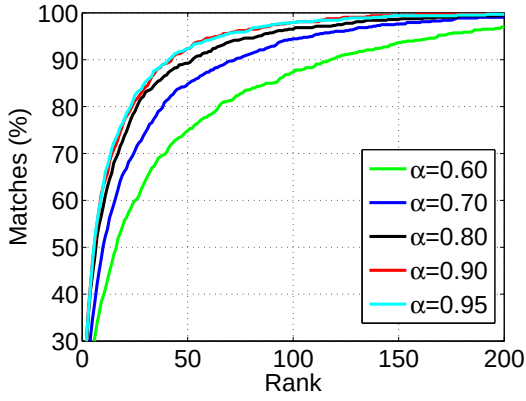


Fig. 6. CMC curves of our method versus varying values of α at $p = 316$ on the VIPeR dataset.

Table 4. Average training time (seconds) of different metric learning methods on the VIPeR dataset.

Method	Time
RBML	0.0020
KISSME	0.0006
RS-KISSME	0.0020
Mahalanobis	0.0002
Identity	0.000001
ITML	6.5100
LMNN	10.2200

from 0 to 1, and $\tau = (1/d)\text{tr}(\Sigma)$. Table 5 shows the matching rates of different regularization methods on the VIPeR dataset. As shown in the table, our method outperforms the smoothing and shrinkage regularization techniques at varying sizes of training set. This indicates that our method is the most effective in estimating the covariance matrix correctly.

Performance of Different Sizes of Training Set: We investigated the performance of our method with varying sizes of training set. As shown in Table 6, our method outperforms other metric learning methods at different training sizes, especially when the size of training set is small.

Table 5. Matching rates (%) of different regularization methods on the VIPeR dataset.

Training Size	$p = 316$				$p = 100$			
Rank	1	5	10	20	1	5	10	20
RBML	**27.53**	**57.6**	**73.3**	**86.4**	**13.5**	**33.6**	**45.7**	**60.0**
Smoothing	21.8	51.9	66.1	77.2	9.9	26.5	38.0	52.4
Shrinkage	26.3	55.9	70.3	84.7	11.1	30.3	41.2	54.2

Table 6. Rank 1 matching rates (%) at varying p on the VIPeR dataset.

Method	p=500	p=300	p=200	p=100
RBML	**43.2**	**28.3**	**19.2**	**13.6**
KISSME	40.9	23.2	15.1	1.3
RSKISSME	41.7	24.4	18.1	10.9
Mahalanobis	20.4	13.3	3.9	3.5

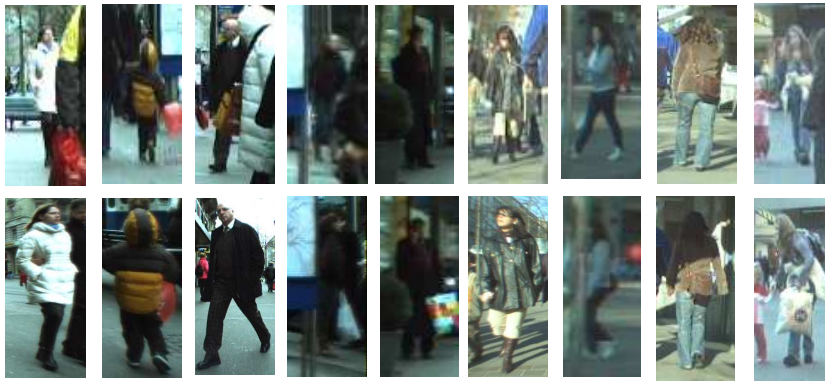


Fig. 7. Sample images of the ETHZ dataset. Each column show the original image of the same person with different pose captured from a moving camera.

3.3 Evaluation on the ETHZ dataset

The ETHZ dataset [29] consists of 8555 images of 146 persons. The images are obtained from 3 video sequences captured from a moving camera. Since the persons are captured in one single camera, the pose and viewpoint variations in this dataset are smaller than those in the VIPeR dataset. However, there are larger variations in illumination and occlusions in this dataset. Fig. 7 shows some samples from the ETHZ dataset. We randomly selected p persons and for each person, we also randomly choose 6 images for training. In the testing phase, we also used the single-shot testing where one image per person was randomly selected to form the gallery set and the remaining

Table 7. Matching rates (%) of different metric learning methods on the ETHZ dataset.

Method	$p = 76$				$p = 26$			
Rank	1	5	10	20	1	5	10	20
RBML	**71.65**	**89.00**	**94.39**	**97.74**	62.30	**82.46**	**88.23**	**93.79**
KISSME [5]	70.93	87.11	93.52	97.31	42.67	72.62	82.85	92.1
RS-KISSME [4]	70.47	88.57	93.81	97.34	60.10	79.40	86.43	92.57
Mahalanobis	71.28	88.38	94.33	97.47	**62.42**	79.42	86.31	93.35
Identity	56.29	79.48	88.18	93.87	51.24	72.07	80.69	88.58
ITML [24]	43.92	70.85	82.09	99.31	53.92	75.63	84.45	91.97
LMNN [23]	65.76	85.78	92.23	96.22	60.32	80.01	87.20	93.02

Fig. 8. Examples of person re-identification on the ETHZ dataset using our proposed RBML.

Fig. 9. Sample images of the i-LIDS MCTS dataset. Each column show the original image of the same person captured from different cameras.

images were used to form the probe set. Table 7 shows the matching rates of different metric learning methods on the ETHZ dataset. As can be shown, our RBML achieves comparable results with existing metric learning when p is set to 76, and outperforms other methods when p is set to 26. Fig. 8 shows the matching examples of our RBML method on the ETHZ dataset.

3.4 Evaluation on the i-LIDS dataset

The i-LIDS Multiple-Camera Tracking Scenario (MCTS) dataset [41] contains 476 images from 119 persons, where each person has an average of 4 images. This dataset was captured at a busy airport arrival hall by using multiple non-overlapping cameras. It is one of the more difficult datasets for person re-identification because there are heavy occlusions caused by the busy crowd and large illumination and pose variations caused by different camera views. Fig. 9 shows some samples in the dataset. We randomly se-

Fig. 10. Examples of person re-identification on the i-LIDS dataset using our proposed RBML.

Table 8. Matching rates (%) of different metric learning methods on the i-LIDS dataset.

Training Size	$p = 89$				$p = 39$			
Rank	1	5	10	20	1	5	10	20
RBML	**32.78**	**63.07**	**81.25**	**94.15**	**20.24**	**44.07**	**55.58**	**72.27**
KISSME [5]	31.24	57.54	75.91	92.84	10.72	26.95	37.10	51.79
RS-KISSME [4]	30.58	61.93	80.11	**94.15**	18.51	39.83	53.46	69.97
Mahalanobis	29.44	50.37	68.13	89.76	12.77	29.86	41.11	53.91
Identity	18.73	46.64	64.76	89.42	13.69	30.60	44.79	58.99
ITML [24]	6.63	25.15	48.17	81.14	6.07	18.42	29.25	49.58
LMNN [23]	23.63	57.18	76.12	92.28	17.79	38.19	51.10	65.45

lected p persons for training and used the single-shot testing for the remaining subjects. Table 8 shows the matching rates of different metric learning methods on the i-LIDS M-CTS dataset. We see that our RBML outperforms other metric learning methods. Fig. 10 shows the matching examples of our RBML method on the i-LIDS dataset.

3.5 Discussion

The above experimental results suggest the following two observations:

1. RBML consistently outperforms the other metric learning methods, especially when the size of the training set is relatively small. That is because the noise space is much larger when the size of the training set is small, and our regularization model plays an important role for such scenarios.
2. Each individual part of our regularization model contributes to the improvement of the identification performance. Moreover, our method obtains better performance than other existing regularization methods. This is because the RBML estimates the eigen-spectrum in a parametric manner so that a stable eigen-spectrum can be obtained.

4 Conclusion

In this paper, we have proposed a regularized Bayesian metric learning (RBML) for person re-identification. The proposed method learns a Mahalanobis distance metric by measuring the probability ratio between similar and dissimilar pairs modeled with a Bayesian model, where the covariance matrices are regularized in a parametric manner so that they are well-estimated. Experimental results on three widely used re-identification datasets have shown the the efficacy of the proposed method.

Acknowledgement

This study is partially supported by the research grant for the Human Cyber Security Systems (HCSS) Program at the Advanced Digital Sciences Center (ADSC) from the Agency for Science, Technology and Research (A*STAR) of Singapore, and the Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 20130191120033).

References

1. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: CVPR. (2010) 2360–2367
2. Kviatkovsky, I., Adam, A., Rivlin, E.: Color invariants for person re-identification. PAMI **35**(7) (2012) 1622–1634
3. Zheng, W., Gong, S., Xiang, T.: Re-identification by relative distance comparison. PAMI **35**(3) (2013) 653–668
4. Tao, D., Jin, L., Wang, Y., Yuan, Y., Li, X.: Person re-identification by regularized smoothing kiss metric learning. TCSVT **23**(10) (2013) 1675–1685
5. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: CVPR. (2012) 2288–2295
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV **60**(2) (2004) 91–110
7. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: ECCV. (2006) 404–417
8. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. PAMI **28**(12) (2006) 2037–2041
9. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: ECCV. (2008) 262–275
10. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: CVPR. (2013) 3586–3593
11. Moghaddam, B., Jebara, T., Pentland, A.: Bayesian face recognition. Pattern Recognition **33**(11) (2000) 1771–1782
12. Lu, J., Zhang, E.: Gait recognition for human identification based on ica and fuzzy svm through multiple views fusion. Pattern Recognition Letters **28**(16) (2007) 2401–2411
13. Lu, J., Tan, Y.: Uncorrelated discriminant simplex analysis for view-invariant gait signal computing. Pattern Recognition Letters **31**(5) (2010) 382–393
14. Lu, J., Tan, Y.P.: Gait-based human age estimation. IEEE Transactions on Information Forensics and Security **5**(4) (2010) 761–770

15. Liu, N., Lu, J., Tan, Y.P.: Joint subspace learning for view-invariant gait recognition. IEEE Signal Processing Letters **18**(7) (2011) 431–434
16. Chen, Y.C., Patel, V.M., Phillips, P.J., Chellappa, R.: Dictionary-based face recognition from video. In: European Conference on Computer Vision. (2012) 766–779
17. Lu, J., Tan, Y.P.: Ordinary preserving manifold analysis for human age and head pose estimation. IEEE Transactions on Human-Machine Systems **43**(2) (2013) 249–258
18. Lu, J., Wang, G., Moulin, P.: Human identity and gender recognition from gait sequences with arbitrary walking directions. IEEE Transactions on Information Forensics and Security **9**(1) (2014) 51–61
19. Lu, J., Wang, G., Moulin, P.: Image set classification using holistic multiple order statistics features and localized multi-kernel metric learning. In: IEEE International Conference on Computer Vision. (2013) 329–336
20. Lu, J., Tan, Y.P., Wang, G.: Discriminative multimanifold analysis for face recognition from a single training sample per person. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(1) (2013) 39–51
21. Lu, J., Zhou, X., Tan, Y.P., Shang, Y., Wang, G.: Neighborhood repulsed metric learning for kinship verification. IEEE Transactions on Pattern Analysis and Machine Intelligence **36**(2) (2014) 331–345
22. Yan, H., Lu, J., Deng, W., Zhou, X.: Discriminative multimetric learning for kinship verification. IEEE Transactions on Information Forensics and Security **9**(7) (2014) 1169–1178
23. Weinberger, K., Saul, L.: Distance metric learning for large margin nearest neighbor classification. JMLR **10** (2009) 207–244
24. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: ICML. (2007) 209–216
25. Globerson, A., Roweis, S.: Metric learning by collapsing classes. In: NIPS. (2005) 451–458
26. Jungling, K., Bodensteiner, C., Arens, M.: Person re-identification in multi-camera networks. In: CVPRW. (2011) 55–61
27. Ma, B., Su, Y., Jurie, F.: Bicov: a novel image representation for person re-identification and face verification. In: BMVC. (2012) 1–6
28. Ma, B., Su, Y., Jurie, F.: Local descriptors encoded by fisher vectors for person re-identification. In: ECCVW. (2012) 413–422
29. Schwartz, W.R., Davis, L.S.: Learning discriminative appearance-based models using partial least squares. In: SIBGRAPI. (2009) 322–329
30. Prosser, B., Zheng, W.S., Gong, S., Xiang, T., Mary, Q.: Person re-identification by support vector ranking. In: BMVC. (2010) 1–11
31. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning. In: ACCV. (2012) 31–44
32. Loy, C.C., Liu, C., Gong, S.: Person re-identification by manifold ranking. In: ICIP. (2013) 3567–3571
33. Mignon, A., Jurie, F.: Pcca: A new approach for distance learning from sparse pairwise constraints. In: CVPR. (2012) 2666–2672
34. Dikmen, M., Akbas, E., Huang, T.S., Ahuja, N.: Pedestrian recognition with a learned metric. In: ACCV. (2011) 501–512
35. Hirzer, M., Roth, P.M., Köstinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: ECCV. (2012) 780–793
36. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: CVPR. (2011) 649–656
37. Jiang, X., Mandal, B., Kot, A.: Eigenfeature regularization and extraction in face recognition. PAMI **30**(3) (2008) 383–394
38. Lu, J., Tan, Y.P.: Regularized locality preserving projections and its extensions for face recognition. TSMC Part B **40**(3) (2010) 958–963

39. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: PETS. (2007)
40. Ess, A., Leibe, B., Van Gool, L.: Depth and appearance for mobile scene analysis. In: ICCV. (2007) 1–8
41. Zheng, W.S., Gong, S., Xiang, T.: Associating groups of people. In: BMVC. (2009) 1–11
42. Cheng, D.S., Cristani, M., Stoppa, M., Bazzani, L., Murino, V.: Custom pictorial structures for re-identification. In: BMVC. (2011) 1–6
43. Liu, C., Gong, S., Loy, C.C., Lin, X.: Person re-identification: what features are important? In: ECCV. (2012) 391–401
44. Friedman, J.H.: Regularized discriminant analysis. Journal of the American Statistical Association **84**(405) (1989) 165–175