Statistical Inference of Gaussian-Laplace Distribution for Person Verification

Zheng Wang
National Engineering Research Center for Multimedia Software, Wuhan University, China
wangzwhu@whu.edu.cn

Ruimin Hu
National Engineering Research Center for Multimedia Software, Wuhan University, China
hrm@whu.edu.cn

Yi Yu
National Institute of Informatics, Japan
yiyu@nii.ac.jp

Junjun Jiang
School of Computer, China University of Geosciences, Wuhan, China
junjun0595@163.com

Jiayi Ma
School of Electronic Information, Wuhan University, China
jyma2010@gmail.com

Shin'ichi Satoh
National Institute of Informatics, Japan
satoh@nii.ac.jp

ABSTRACT

Metric learning is an important issue in the person verification problem, which is to identify whether a pair of face or human body images is about the same person. Due to low running cost, the non-iterative statistical inference methods for metric learning show their efficiency and effectiveness to large scale datasets and on-line updating person verification applications. The KISSME method is a typical one that constructs the metric based on two assumptions that both of the discrepancy spaces of negative pairs and positive pairs should be Gaussian structures. However, we find that, in fact, the distribution of discrepancies of positive pairs might tend to the Laplace distribution rather than the Gaussian distribution. Based on this finding, we propose a metric learning method by exploiting Gaussian-Laplace distribution statistical inference, where the Gaussian distribution of negative discrepancies and the Laplace distribution of positive discrepancies are considered together. Experiments conducted on two human body datasets (VIPeR and Market-1501) and one face dataset (LFW) show its superiority in terms of effectiveness and efficiency as compared with the state-of-the-art approaches, no matter the appearance description is handcrafted or deep learned.

CCS CONCEPTS

• **Information systems** → **Image search**; *Probabilistic retrieval models*; • **General and reference** → *Metrics*; *Verification*;

MM'17, October 23–27, 2017, Mountain View, CA, USA.
 ISBN 978-1-4503-4906-2/17/10...$15.00
DOI: http://dx.doi.org/10.1145/3123266.3123421

KEYWORDS

Metric Learning, Person Verification, Statistical Inference

1 INTRODUCTION

Person verification trying to answer the question "Are you the person you claim to be", is an important task, which can be exploited in a lot of critical applications [16], such as long-term multi-camera tracking [13], forensic search [28] and video retrieval systems [43]. The task is to verify whether photos contain the same person, or re-identify persons in a huge amount of surveillance cameras [14, 30–32, 35, 36], even if the person is not seen before. There are two main visual clues for person verification: face images and human body figures. It is well known that the probe image and the other images in the gallery are rarely collected in the same environment, which leads to large intra-person variations including resolution differences, illumination changes, and viewpoint transformations. As a result, person verification is a very challenging task.

To overcome the difficulties in person verification, previous research efforts primarily focused on two aspects. One is appearance representation [1, 5, 6, 17–20, 33, 34, 40], which aims at constructing discriminative visual descriptors to represent the face or the human body. However, designing a robust feature description which adapts to different realistic conditions is challenging [44]. The other is distance metric [12, 16, 17, 27, 44], which tries to learn a proper distance metric function to calculate the distance of the former descriptions. This paper focuses on the latter aspect, which aims at designing an effective distance metric function for person verification. Even though excellent visual descriptions could be found in the future, we believe that a well learned metric function could still promote the results. Considering formulation properties and optimization processes, we divide those distance metric methods into two categories:

Iterative optimization. Most of the approaches are based on the class of distance functions with an iterative optimization scheme, such as Probabilistic Relative Distance Comparison (PRDC) [44], Locally-Adaptive Decision Functions

(LADF) [16], Cross-view Quadratic Discriminant Analysis (XQDA) [17], Multi-Kernel Metric Learning (MKML) [22], Similarity Learning with Spatial Constraints (SCSP) [2]. These methods pay attention to designing loss functions and regularizing solutions by learning prior knowledge from abundant labeled samples. However, almost all the algorithms are required to iteratively learn the optimal parameters of their designed metric functions. Generally, repeated iterations are computationally expensive especially for large scale datasets. Hence, this kind of approaches are not proper when labeled samples are increasing or a situation needs on-line training.

Non-iterative statistical inference. Another type of approaches tries to obtain the metric function with a non-iterative optimization way, such as Mahalanobis distance, KISSME [12], RS-KISSME [24], DR-KISSME [23]. Motivated by a statistical inference perspective, the KISSME based methods consider the generation process for observed sample pairs. After the distribution of sample pairs is discovered, given a new pair $(\mathbf{x}_i, \mathbf{x}_j)$, its impossibility of belonging to the same person, $\delta(\mathbf{x}_i, \mathbf{x}_j)$, will be calculated by the likelihood of pair discrepancies. $\delta(\mathbf{x}_i, \mathbf{x}_j)$ essentially represents the distance. Without an iteration process, it just involves computation of distributional parameters before verification, so it is scalable to large datasets and on-line tasks.

Nowadays, the explosion of data asks for a scalable distance metric function, which can quickly learn a new model and adapt to the data captured from new environments. Once more labeled data are poured in, the models should be quickly updated with the incremental data rather than re-trained for a long waiting time. In this situation, we consider that a non-iterative statistical inference metric learning method might be a better choice, which learns the model by non-iterative parameter calculation operation. Among those non-iterative methods, the KISSME based methods show their outstanding effectiveness [12, 42]. To construct the distance function, the KISSME method decides whether a feature pair $(\mathbf{x}_i, \mathbf{x}_j)$ is negative or positive by a likelihood ratio test as Eq. 1.

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \frac{p(\mathbf{x}_{ij}, H_-)}{p(\mathbf{x}_{ij}, H_+)}. \tag{1}$$

The formulation includes two independent generation processes respectively for observed commonalities of negative (the numerator part) and positive (the denominator part) pairs, where $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ denotes the discrepancy of features, and H_- and H_+ are two hypothesises in the space of pairwise discrepancies. It assumes that both discrepancy spaces of negative pairs and positive pairs should be Gaussian structures.

However, we consider that the distributions of positive discrepancies (intra-person variations) and negative discrepancies (inter-person variations) should be different. The inter-person variations are often large, hence the values of the discrepancies are dense and random. Whereas, the intra-person variations are often small, hence the values of the discrepancies are sparse. Discrepancies of positive pairs are more likely to have an empirical distribution that is highly peaked at the mean vector and then asymptotically falls off more slowly than the Gaussian distribution as the distance from the mean vector increases. So the discrepancy space of positive pairs is more likely to be the Laplace structure rather than the Gaussian structure. This is verified by Fig. 2, and we investigate the characteristics of the negative and positive discrepancy space in Sec. 2. Considering that positive and negative discrepancy spaces respectively follow a *Laplace* distribution and a *Gaussian* distribution, we construct the proposed distance metric function by statistical inference, which shows both considerable effectiveness and efficiency (as Fig.1 illustrates). The contributions of this paper are as follows:

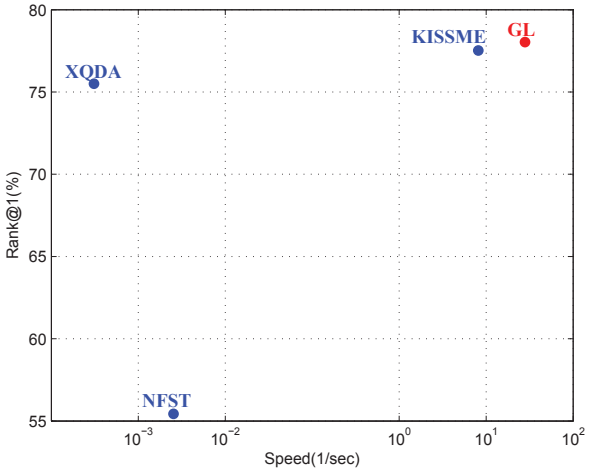

Figure 1: Training speed vs. Accuracy at Rank 1 for the tested methods on the Market-1501 dataset. XQDA [17] and NFST [39] are two iterative optimization based metric learning methods, and KISSME [12] is a non-iterative statistical inference based metric learning method, which are shown in blue fond. Our GL method (shown in red) provide both high speed and accuracy. More details are listed in Sec. 4.

(1) Observing the distributions of discrepancies through preliminary experiments, we find the phenomena that the distribution of the discrepancy space of negative pairs approximates to the Gaussian structure, while that of positive pairs tends to the Laplace structure.

(2) Based on the above observations, we propose a metric learning method exploiting statistical inference of Gaussian-Laplace distribution (GL). Experiments conduct on three public datasets have validated the effectiveness of the proposed method, with a considerable improvement over state-of-the-art methods with iterative and non-iterative learning.

(3) The proposed method is proved to be very simple and in low running cost, which means that the non-iterative learned distance metric function is more suitable for the large scale and on-line person verification application.

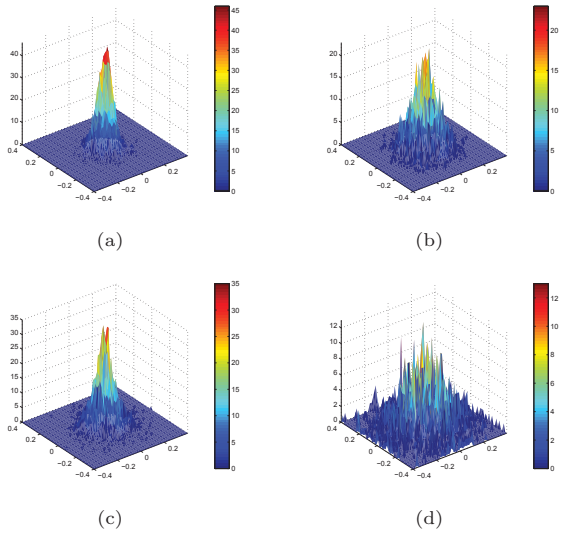

Figure 2: The distribution maps of positive discrepancies and negative discrepancies in 2D space. (a) The distribution map of randomly selected 2D positive discrepancies. (b) The distribution map of randomly selected 2D negative discrepancies. (c) The distribution map of PCA 2D positive discrepancies. (d) The distribution map of PCA 2D negative discrepancies.

2 INVESTIGATION ON DISCREPANCY DISTRIBUTIONS

Intuitively, to construct the distance metric through a non-iterative statistical inference way, we attempt to comprehend the discrepancy distributions of positive pairs and negative pairs. To investigate the issue, we not only **perceptually** visualized the distributions of discrepancies, but also **rationally** took goodness of fit test.

For the perceptual part, several preliminary visualization experiments were made to observe the distribution structure. 3000 positive and 3000 negative image pairs were obtained from the LFW dataset [11]. For each image, a 256 dimension feature was extracted by the Lightened CNN model trained on MS-Celeb-1M by caffe-rc3 [33]. Then, for each image pair, a discrepancy was computed by the difference of the two corresponding image features. Hence, we obtained 3000 discrepancies of positive pairs and 3000 discrepancies of negative pairs. To respectively understand the distributions of positive and negative discrepancies, we visualized these two distributions in 2D space. In order to get an comparatively overall observation of the distributions of discrepancies in different dimensions, we selected typical dimensions by random components and principle components.

First, we randomly selected two dimensions to view the general distribution map of discrepancies. Fig. 2(a) and Fig. 2(b) respectively show the distribution maps of positive discrepancies and negative discrepancies. As can be seen from the shape of these two maps, the distribution of positive discrepancies does not look like the previous assumed Gaussian distribution. The distribution, prone to the Laplace distribution, is more highly peaked at the center than that of negative discrepancies, extremely following the Gaussian distribution. Second, we chose two key dimensions by Principal Component Analysis (PCA) to view the primary distribution map of discrepancies. Fig. 2(c) and Fig. 2(d) respectively show the distribution maps of positive discrepancies and negative discrepancies. We can discover the same phenomenon as the former general distribution maps demonstrate. Comparing the map shapes under these two different dimension selection methods, we can also find that the distributions generated by PCA discrepancies with principle dimensions look more similar to Laplace or Gaussian distributions.

It should be clearly noted that the distribution maps demonstrate similar shapes as described above, even if we exploit different feature descriptors on different datasets. We do not fully explore the distributions of all the datasets, but take the Lightened CNN model and LFW dataset as representative, because the feature dimensions and the dataset volume are moderate.

For the rational part, to prove this observation, we take goodness of fit test to different distributions by utilizing the tool Minitab[1]. We continually selected each corresponding dimension of all the discrepancies, and study their distributions with Minitab.

Tab. 1, Tab. 2, Tab. 3, and Tab. 4 respectively show the values of goodness of fit test on the distributions of negative discrepancies and positive discrepancies with one dimension and all dimensions in average. The tables include Anderson-Darling (AD) statistics and the corresponding P-value for each distribution. The AD statistic measures how well the data follow a particular distribution. The better the distribution fits the data, the smaller this statistic will be. In addition, if the P-value (when available) for the AD test is higher than the chosen significance level (usually 0.05), it will conclude that the data follow the specified distribution. We can see that $P > 0.05$ for the Gaussian distribution in Tab. 1 and Tab. 3, which means that the distribution of negative discrepancies fits a Gaussian very well. We can also see that $AD = 8.658$ and $AD = 7.158$ for the Exponential distribution are respectively the lowest value in Tab. 2 and Tab. 4, which means that the half part of the distribution of positive discrepancies is more likely to be an Exponential, compared with the other distributions. Since a Laplace distribution can be thought of as two Exponential distributions spliced together back-to-back, we consider the whole

[1]Minitab is a powerful statistical software, which is available at http://www.minitab.com/. The tool can take goodness of fit test to different distributions, such as Gaussian, Exponential, Smallest Extreme Value, Largest Extreme Value and Logistic. Since no Laplace test exists in the tool, we insteadly focus on whether the positive value part follows Exponential distribution.

distribution of positive discrepancies is more likely to be a Laplace distribution.

Based on the above analysis, to better verify persons, we propose to learn a distance metric function by statistical inference considering not only Gaussian distribution of negative discrepancies but also Laplace distribution of positive discrepancies.

Table 1: Goodness of fit test on the distribution of negative discrepancies with one dimension.

Distribution	*AD*	*P*
Gaussian	**0.658**	**0.086**
Exponential	714.426	< 0.01
Smallest Extreme Value	51.418	< 0.01
Largest Extreme Value	38.363	< 0.01
Logistic	1.348	< 0.005

Table 2: Goodness of fit test on the distribution of positive discrepancies with one dimension (only use the positive value part of discrepancy).

Distribution	*AD*	*P*
Gaussian	48.657	< 0.05
Exponential	**8.658**	**<0.05**
Smallest Extreme Value	94.172	< 0.010
Largest Extreme Value	10.241	< 0.010
Logistic	25.178	< 0.005

Table 3: Goodness of fit test on the distribution of negative discrepancies with all dimensions in average.

Distribution	*AD*	*P*
Gaussian	**0.612**	**0.092**
Exponential	701.124	< 0.01
Smallest Extreme Value	53.312	< 0.01
Largest Extreme Value	40.112	< 0.01
Logistic	1.945	< 0.005

Table 4: Goodness of fit test on the distribution of positive discrepancies with all dimensions in average (only use the positive value part of discrepancy).

Distribution	*AD*	*P*
Gaussian	49.287	< 0.05
Exponential	**7.158**	**<0.05**
Smallest Extreme Value	98.123	< 0.010
Largest Extreme Value	13.932	< 0.010
Logistic	26.291	< 0.005

3 THE PROPOSED METHOD

Our method considers two independent generation processes for observed samples of positive and negative pairs. The impossibility is defined by the possibility of belonging to either one or the other. From a statistical inference point of view the optimal statistical decision on whether a pair $(\mathbf{x}_i, \mathbf{x}_j)$ is positive or not can be obtained by a likelihood ratio test. Thus, we test the hypothesis H_- that a pair is negative ($y_{ij} = 0$) versus the hypothesis H_+ that the pair is positive ($y_{ij} = 1$), following the KISSME method as Eq. 1. A high value of $\delta(\mathbf{x}_i, \mathbf{x}_j)$ means that H_- is recommended and the pair is considered as negative, in contrast, a low value means that H_- is rejected and the pair is considered as positive. To facilitate the discussion, we make the following definitions:

$$\Sigma_- = \frac{1}{N} \sum_{y_{ij}=0} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top, \tag{2}$$

$$\Sigma_+ = \frac{1}{N} \sum_{y_{ij}=1} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top, \tag{3}$$

$$\phi(\mathbf{x}_{ij}, \Sigma_-) = (\mathbf{x}_i - \mathbf{x}_j)^\top \Sigma_-^{-1} (\mathbf{x}_i - \mathbf{x}_j), \tag{4}$$

$$\phi(\mathbf{x}_{ij}, \Sigma_+) = (\mathbf{x}_i - \mathbf{x}_j)^\top \Sigma_+^{-1} (\mathbf{x}_i - \mathbf{x}_j). \tag{5}$$

In these definitions, N denotes the number of positive or negative training pairs. Σ_+^{-1} and Σ_-^{-1} stand for the inversions of Σ_+ and Σ_-. Based on the above discussion, we reform Eq. 1 to the following form:

$$\delta(\mathbf{x}_i, \mathbf{x}_j) = \frac{p(\mathbf{x}_{ij}, H_-)}{p(\mathbf{x}_{ij}, H_+)} = \frac{f_G(\mathbf{x}_{ij}|\Sigma_-)}{f_L(\mathbf{x}_{ij}|\Sigma_+)}. \tag{6}$$

Here, $f_G(\mathbf{x}_{ij}|\Sigma_-)$ is a PDF (Probability Density Function) with the parameter Σ_- for hypothesis H_- that the sample pair $(\mathbf{x}_i, \mathbf{x}_j)$ is negative. As discussed in Sec. 2, the discrepancy space of negative pairs follows a Gaussian structure (Eq. 7). Meanwhile, $f_L(\mathbf{x}_{ij}|\Sigma_+)$ is a PDF with the parameter Σ_+ for hypothesis H_+ that the sample pair $(\mathbf{x}_i, \mathbf{x}_j)$ is positive, where the discrepancy space of positive pairs follows a Laplace structure (Eq. 8, the formulation refers to [4]).

$$f_G(\mathbf{x}_{ij}|\Sigma_-) = \frac{\exp(-\frac{1}{2}\phi(\mathbf{x}_{ij}, \Sigma_-))}{\sqrt{2\pi\Sigma_-}}, \tag{7}$$

and

$$f_L(\mathbf{x}_{ij}|\Sigma_+) = \frac{1}{(2\pi)^{(d/2)}} \frac{2}{\lambda} \frac{K_{(d/2)-1}(\sqrt{\frac{2}{\lambda}\phi(\mathbf{x}_{ij}, \Sigma_+)})}{(\sqrt{\frac{\lambda}{2}\phi(\mathbf{x}_{ij}, \Sigma_+)})^{(d/2)-1}}, \tag{8}$$

where $K_m(x)$ denotes the modified Bessel function of the second kind and order m, evaluated at x. d denotes the dimension of sample feature, and $\lambda > 0$ stands for a scale parameter.

By substituting Eq. 7 and Eq. 8 into Eq. 6, and taking the log operation, Eq. 6 can be transformed to Eq. 9, where the terms independent to $(\mathbf{x}_i, \mathbf{x}_j)$, such as the denominator

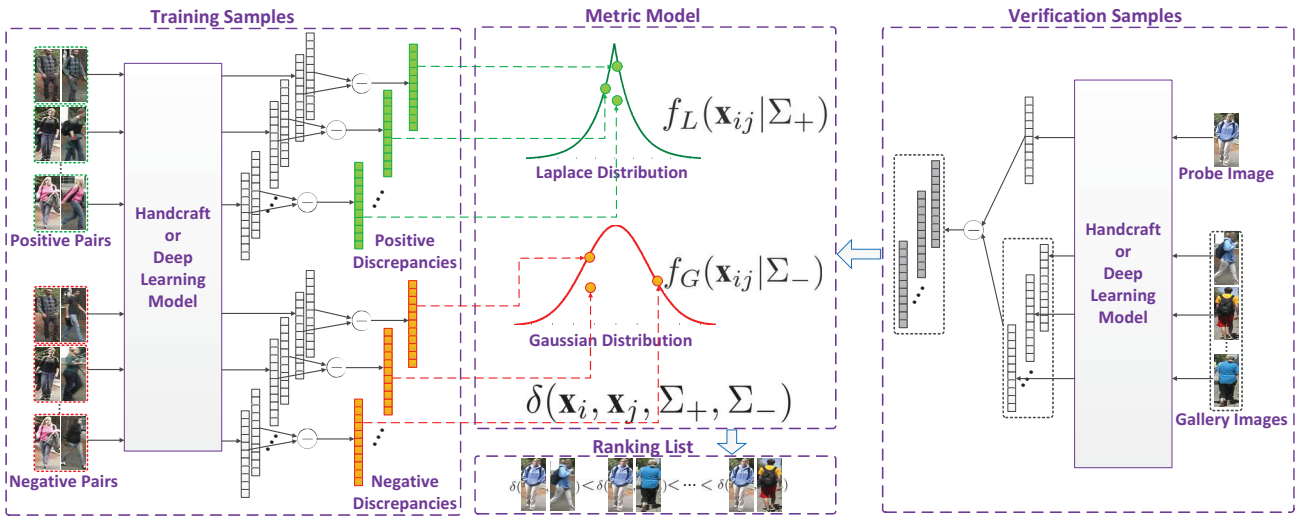

Figure 3: The framework of the proposed method. First, training samples generate positive and negative discrepancies. Second, following two different distributions, the parameters Σ_+ and Σ_- are calculated to form the model of distance metric function. Third, this model is evaluated with verification samples.

term of Eq. 7, is omitted.

$$\delta(\mathbf{x}_i, \mathbf{x}_j) \sim \log\left(\frac{\exp(-\frac{1}{2}\phi(\mathbf{x}_{ij}, \Sigma_-))(\sqrt{\frac{\lambda}{2}\phi(\mathbf{x}_{ij}, \Sigma_+)})^{(d/2)-1}}{K_{(d/2)-1}(\sqrt{\frac{2}{\lambda}\phi(\mathbf{x}_{ij}, \Sigma_+)})}\right). \tag{9}$$

Further, we strip constant terms as they just provide an offset and Eq. 9 is simplified to

$$\begin{aligned}\delta(\mathbf{x}_i, \mathbf{x}_j, \Sigma_+, \Sigma_-) = &(\frac{d-2}{2})\log(\sqrt{\frac{\lambda}{2}\phi(\mathbf{x}_{ij}, \Sigma_+)}) \\ &- \log(K_{(d/2)-1}(\sqrt{\frac{2}{\lambda}\phi(\mathbf{x}_{ij}, \Sigma_+)})) \\ &- \frac{1}{2}\phi(\mathbf{x}_{ij}, \Sigma_-).\end{aligned} \tag{10}$$

This equation essentially defines the distance metric function, whose metric characteristic is decided by the parameters Σ_+ and Σ_-. Hence, the whole process of the proposed method can be described as follows (as Fig. 3 illustrates). First, features are extracted from training samples by using a handcraft or deep learning model. Positive pairs generate positive discrepancies, and negative pairs generate negative discrepancies. Second, the parameters Σ_+ and Σ_- are calculated to construct the model of distance metric function. Third, this model is evaluated with verification samples. The algorithm is demonstrated in Alg. 1.

In particular, as Eq. 2 and Eq. 3 demonstrate, the parameters Σ_+ and Σ_- are calculated by the sum operation on the outer product of pairwise discrepancies. It means that if new labeled data are added, we only need to calculate the parameters Σ'_+ and Σ'_-, and directly add them to the initial parameters Σ_+ and Σ_-. The model is very easy for the incremental updating, hence the method is suitable for the on-line application.

Algorithm 1 Algorithm of the proposed method.

Input: training samples which include positive pairs $\{(\mathbf{x}_i, \mathbf{x}_j)|y_{ij} = 1\}$ and negative pairs $\{(\mathbf{x}_i, \mathbf{x}_j)|y_{ij} = 0\}$, and verification samples which include the probe image $\mathbf{x}_p$ and the gallery images $\{\mathbf{x}_g\}$.

Output: a ranking list of $\{\mathbf{x}_g\}$ for the probe image $\mathbf{x}_p$.

1: calculate the matrix Σ_+ by Eq. 3 with training positive pairs;
2: calculate the matrix Σ_- by Eq. 2 with training negative pairs;
3: **for** each g in the gallery **do**
4: compute the distance $\delta(\mathbf{x}_p, \mathbf{x}_g, \Sigma_+, \Sigma_-)$ between $\mathbf{x}_p$ and $\mathbf{x}_g$ by Eq. 10;
5: **end for**
6: rank the distances from low to high and generate the ranking list.

Complexity Analysis: First, the proposed method learns the metric by non-iterative statistical inference, rather than iterative optimization. It is well known that the algorithm of training iterative optimization based metrics are more complicated. So we can find that the iterative optimization based metric learning methods cost more than $10,000$ times training time compared with the proposed method [12]. Second, the parameters are calculated by statistical inference, and the algorithm complexity is $O(N)$ for calculating Σ_+ and Σ_-, where N stands for the number of image pairs. As we know, for the KISSME method, the calculated Σ_+ and Σ_- should be combined together to form a new Mahalanobis-like

metric. Whereas for the proposed method, the two parameters are used separately, and do not need to be fused into one. Hence, the proposed method is the most efficient metric learning method for on-line training in person verification. Third, the testing algorithm complexity is $O(M)$, where M is the number of gallery images. However, it is reported that only $2.5ms$ is cost for computing each pair of images, which is pretty sufficient for real-time applications.

4 EXPERIMENTS

To evaluate the applicability of the proposed method, we conduct experiments on various standard benchmarks with rather diverse characteristics. We first conduct experiments on human body figures on the widely used dataset VIPeR in Sec. 4.1. Then, in Sec. 4.2, we expand the evaluation on the largest dataset Market-1501. In Sec. 4.3, we study the problem on faces in unconstrained environments. Finally, the running cost of the proposed method is evaluated and compared.

4.1 VIPeR dataset

The widely used VIPeR dataset [7] contains $1,264$ outdoor images obtained from two views of 632 persons. Each person has a pair of images taken from two different cameras respectively. All images of individuals are normalized to a size of 128×48 pixels. View changing, illumination and image quality variations are the causes of appearance change. Some example images are shown in Fig. 4(a). We respectively list some positive pairs, where two images of each pair are from the same identity, and some negative pairs from different identities. We followed the general evaluation protocol [12], and split the set of 632 image pairs randomly into two sets of 316 image pairs each, one for training and one for testing, and computed the average over 5 runs. To indicate the performance of the various algorithms, we report Cumulative Matching Characteristic (CMC) [29] at different ranks. CMC represents the expectation of the true match being found within the first n ranks.

To prove that the proposed method (GL) can improve the effectiveness of different feature representations, no matter handcraft or deep learning, we conducted GL on two handcraft feature methods SCNCD [34] and LOMO [17], and one deep learning feature method FTCNN [20]. Meanwhile, these features were exploited and evaluated on the iterative optimization metric learning method XQDA [17] and the non-iterative statistical inference metric learning method KISSME [12]. Tab. 5 shows the results at different ranks. From the bottom part of the table, we can find that the proposed GL **surpasses the KISSME with all the three features**, surpasses the XQDA with the FTCNN feature, and performs not so well as the XQDA with the LOMO feature at the first 10 ranks. In particular, exploiting FTCNN feature, the proposed method obtains a nearly 10% promotion at rank one. We consider the reason is that the XQDA metric is specially designed for LOMO in [17]. It should be mentioned that we do not list the results of SCNCD+XQDA, because we have not seen any work combining this two methods. Hence, we believe that the proposed metric is effective for different features on the VIPeR dataset.

Table 5: Comparing results with the-state-of-the-art person verification methods on top ranked matching rate (%) on VIPeR dataset. We utilize the red font to stand for the highest performance among all the list methods, and the blue font to denote the second highest performance. The rules are the same for the following two tables.

Method(rank@)	1	5	10	20
PRDC	15.7	38.4	53.9	70.1
SDALF	19.9	38.4	49.4	66
BiCov	20.6	43.2	56.1	68
eSDC	26.3	46.4	58.6	72.8
DeepMetric	28.2	59.3	73.4	86.4
LADF	30	64	80	**92**
CPDL	34	64.2	77.5	88.6
MKML	37	**69.9**	80.7	90.1
DeepFeature	**40.5**	60.8	70.4	84.4
SCNCD+KISSME	34.81	66.77	80.38	89.56
SCNCD+GL	38.92	**69.3**	**81.33**	90.51
LOMO+XQDA	40	68.13	80.51	91.08
LOMO+KISSME	28.48	57.59	76.58	89.24
LOMO+GL	35.76	65.19	78.16	91.14
FTCNN+XQDA	31.2	59.8	74	83.5
FTCNN+KISSME	31.53	62.86	78.37	88.18
FTCNN+GL	**40.39**	67.97	**81.9**	**91.66**

In addition, we compared the results with the state-of-the-art methods, which are PRDC [44], SDALF [5], BiCov [18], eSDC [40], DeepMetric [37], LADF [16], CPDL [15], MKML [22], and DeepFeature [3]. As can be seen, most of the good performances are achieved by exploiting the proprosed GL.

4.2 Market-1501 dataset

The Market-1501 dataset [41] contains 32668 annotated bounding boxes of 1501 identities. Images of each identity are captured by at most six cameras. As far as we know, it is the largest person re-identification dataset to date. Some example images are shown in Fig. 4(b). We followed the general evaluation protocol on this dataset [41]. To indicate the performance of the various algorithms, we report not only CMC values at different ranks, but also the mAP (mean Average Precision [41]) value as described in [41]. We evaluate the mAP values for the Market-1501 dataset, because multiple target images exist in the gallery. Both of these two evaluation criterions demonstrate the effectiveness of different methods.

To prove the effectiveness, similar to evaluations on the VIPeR dataset, we conducted GL on the CaffeNet feature and the ResNet-50 feature, which were proved to be the most helpful representations for the Market-1501 dataset. Both of

(a) VIPeR dataset (b) Market-1501 dataset (c) LFW dataset

Figure 4: Some typical samples of three public datasets. The three public datasets are respectively the human body figures datasets (a) VIPeR and (b) Market-1501, and the face images dataset (c) LFW. For each dataset, we list some positive pairs and negative pairs.

Table 6: Comparing results with the-state-of-the-art person verification methods on top ranked matching rate (%) on Market-1501 dataset.

Method(rank@)	1	10	50	mAP
BiCov	8.28	-	-	2.23
LOMO	26.07	-	-	7.75
BoW	35.84	60.33	75.8	14.75
Siamese CNN	65.88	-	-	39.55
LSTM	61.6	-	-	35.3
SCSP	51.9	-	-	26.35
LOMO+NFST	55.43	-	-	29.87
CaffeNet + L2	59.53	85.51	94.39	32.85
CaffeNet + XQDA	61.34	87.26	95.13	37.59
CaffeNet + KISSME	61.46	86.55	94.8	36.63
CaffeNet + GL	62.08	86.82	95.01	36.86
ResNet-50 + L2	75.62	91.89	96.97	50.68
ResNet-50 + XQDA	75.5	91.66	96.94	52.87
ResNet-50 + KISSME	**77.52**	**92.93**	**97.48**	**53.87**
ResNet-50 + GL	**78.03**	**93.44**	**97.68**	**53.76**

these features were designed and trained by [42]. Meanwhile, these features were exploited and evaluated on the basic Euclidean metric (L2), the iterative optimization metric learning method XQDA [17] and the non-iterative statistical inference metric learning method KISSME [12]. Tab. 6 shows the CMC results at different ranks and the mAP results. From the bottom part of the table, we can find that the GL **surpasses the L2, KISSME, XQDA with both features** at the list ranks, and surpasses the KISSME and the XQDA with the CaffeNet feature at mAP value. Hence, we believe that the proposed metric is effective for the largest human body figures dataset.

In addition, we compared the results with the state-of-the-art methods, which are BiCov [18], LOMO [17], BoW [41], Siamese CNN [25], LSTM [26], SCSP [2], and LOMO+NFST [39]. As can be seen from Tab. 6, most of the best performances are achieved by exploiting the GL metric function with the ResNet-50 feature.

4.3 LFW dataset

The LFW dataset [11] contains more than 13000 face images of 5749 subjects collected from the web with large variations in expression, pose, age, illumination, resolution, and so on. We test our algorithm under the standard "image restricted" setting that is particularly designed for verification. In this setting, the dataset is divided into 10 fully independent folds, and it is ensured that not the same person appears across different folds. 300 positive and 300 negative image pairs are provided within each fold. Each time we learn our distance metirc on 9 training folds, and evaluate on the remaining fold. Following [16], pairwise classification accuracy averaged over 10 runs is reported in general [11].

To prove the effectiveness, we conducted the proposed GL and the compared KISSME on the Lightened CNN feature representation with three different models. The three different deep learning models are described in [33]. Tab. 7 shows the classification accuracy results. From the bottom part of the table, we can find that the GL surpasses the KISSME with all the three kinds of features. Hence, we believe that the proposed metric is effective for the face image verification dataset.

In addition, we compared the results with the state-of-the-art methods, which are MERL+Nowak [9], LBP + C-SML [21], DML-eig combined [38], Convolutional DBN [10], LADF [16], DDML [8], and HPEN + HD-Gabor + DDML [45]. As can be seen from Tab. 7, the best performance

Table 7: Comparing results with the-state-of-the-art person verification methods on the LFW dataset under the standard "image restricted".

Method	Accuracy (%)
MERL+Nowak	76.2
LBP + CSML	85.6
DML-eig combined	85.7
Convolutional DBN	87.8
LADF	89.6
DDML	90.7
HPEN + HD-Gabor + DDML	92.8
Lightened CNN (Model A)+KISSME	92.8
Lightened CNN (Model A)+GL	93.2
Lightened CNN (Model B)+KISSME	94.2
Lightened CNN (Model B)+GL	94.4
Lightened CNN (Model C)+KISSME	**98.2**
Lightened CNN (Model C)+GL	**98.4**

is achieved by exploiting the GL metric with the Lightened CNN (Model C) feature. In particular, it is known that the accuracy above 98% would be very high. Hence, we believe that 0.1% promotion in the LFW dataset is challenging.

From the above evaluations on three typical datasets, we conclude that the proposed method can make a considerable improvement through introducing the Laplace distribution, compared with the KISSME method.

4.4 Running cost

We compared the run time of our method with XQDA, NFST and KISSME on the Market-1501 dataset. The Market-1501 dataset are chosen to report the running time, because the sample size of this dataset is the biggest. We calculate the overall training time over 12,936 samples and test time over 3,368 queries. All algorithms are implemented in Matlab (CPU: Intel Core i3-4030U). Tab. 8 shows that for training, our method is the most efficient. Considering the testing time over 3,368 queries, 2.5 ms for each query. The cost is at the same level as the other methods, and is pretty sufficient for real-time applications.

Table 8: Running time comparison on the Market-1501 dataset (in seconds).

	XQDA	NFST	KISSME	GL (Ours)
Training	3233.8	393.1	0.123	**0.0357**
Testing	31.3	1.6	1.387	8.516

5 CONCLUSION

Based on the observations on distributions of discrepancies, we proposed a noniterative metric learning method exploiting statistical inference of Gaussian-Laplace distribution. The method is very easy and fast, and suitable for the on-line application. We evaluate the method on three person verification datasets, and achieve the state-of-the-art results.

ACKNOWLEDGMENTS

The research was supported by the National Nature Science Foundation of China (61231015, U1611461, U1404618, 61562048, 61503288) and the National High Technology Research and Development Program of China (2015AA016306).

REFERENCES

[1] Le An, Mehran Kafai, Songfan Yang, and Bir Bhanu. 2016. Person reidentification with reference descriptor. *IEEE Trans. Circuits Syst. Video Technol.* 26, 4 (2016), 776–787.
[2] Dapeng Chen, Zejian Yuan, Badong Chen, and Nanning Zheng. 2016. Similarity Learning with Spatial Constraints for Person Re-identification. In *CVPR*.
[3] Shengyong Ding, Liang Lin, Guangrun Wang, and Hongyang Chao. 2015. Deep feature learning with relative distance comparison for person re-identification. *Pattern Recogn.* (2015).
[4] Torbjrn Eltoft, Taesu Kim, and Te Won Lee. 2006. On the multivariate Laplace distribution. *IEEE Signal Proc. Let.* (2006).
[5] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. 2010. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*.
[6] Yuan Gao, Jiayi Ma, and Alan L Yuille. 2017. Semi-Supervised Sparse Representation Based Classification for Face Recognition With Insufficient Labeled Samples. *IEEE Trans. Image Process.* 26, 5 (2017), 2545–2560.
[7] Douglas Gray, Shane Brennan, and Hai Tao. 2007. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*.
[8] Junlin Hu, Jiwen Lu, and Yap Peng Tan. 2014. Discriminative Deep Metric Learning for Face Verification in the Wild. In *CVPR*.
[9] Gary B. Huang, Michael J. Jones, and Eric Learned-Miller. 2008. LFW results using a combined Nowak plus MERL recognizer. In *ECCV workshop*.
[10] G. B. Huang, Honglak Lee, and E Learned-Miller. 2012. Learning hierarchical representations for face verification with convolutional deep belief networks. In *CVPR*.
[11] Gary B. Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. 2007. Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments. *Technical Report 07-49, University of Massachusetts* (2007).
[12] Martin Köstinger, Martin Hirzer, Paul Wohlhart, and Peter M. Roth. 2012. Large scale metric learning from equivalence constraints. In *CVPR*.
[13] Xiangyuan Lan, Andy J Ma, Pong C Yuen, and Rama Chellappa. 2015. Joint sparse representation and robust feature-level fusion for multi-cue visual tracking. *IEEE Trans. Image Process.* 24, 12 (2015), 5826–5841.
[14] Qingming Leng, Ruimin Hu, Chao Liang, Yimin Wang, and Jun Chen. 2015. Person re-identification with content and context re-ranking. *Multimed. Tools Appl.* 74, 17 (2015), 6989–7014.
[15] Sheng Li, Ming Shao, and Yun Fu. 2015. Cross-View Projective Dictionary Learning for Person Re-identification. In *IJCAI*.
[16] Zhen Li, Shiyu Chang, Feng Liang, Thomas S Huang, Liangliang Cao, and John R Smith. 2013. Learning Locally-Adaptive Decision Functions for Person Verification. In *CVPR*.
[17] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z. Li. 2015. Person re-identification by Local Maximal Occurrence representation and metric learning. In *CVPR*.
[18] Bingpeng Ma, Yu Su, and Frdric Jurie. 2014. Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image Vision Comput.* (2014).
[19] Tetsu Matsukawa, Takahiro Okabe, Einoshin Suzuki, and Yoichi Sato. 2016. Hierarchical Gaussian Descriptor for Person Re-identification. In *CVPR*.
[20] Tetsu Matsukawa and Einoshin Suzuki. 2016. Person re-identification using cnn features learned from combination of attributes. In *ICPR*.

[21] Hieu V. Nguyen and Li Bai. 2010. Cosine Similarity Metric Learning for Face Verification. In *ACCV*.
[22] Muhammad Adnan Syed and Jianbin Jiao. 2016. Multi-kernel metric learning for person re-identification. In *ICIP*.
[23] D. Tao, Y. Guo, M. Song, Y. Li, Z. Yu, and Y. Y. Tang. 2016. Person Re-Identification by Dual-Regularized KISS Metric Learning. *IEEE Trans. Image Process.* 25, 6 (2016), 2726–2738.
[24] Dapeng Tao, Lianwen Jin, Yongfei Wang, and Yuan Yuan. 2013. Person Re-Identification by Regularized Smoothing KISS Metric Learning. *IEEE Trans. Circuits Syst. Video Technol.* 23, 10 (2013), 1675–1685.
[25] Rahul Rama Varior, Mrinal Haloi, and Gang Wang. 2016. Gated Siamese Convolutional Neural Network Architecture for Human Re-Identification. In *ECCV*.
[26] Rahul Rama Varior, Bing Shuai, Jiwen Lu, Dong Xu, and Gang Wang. 2016. A Siamese Long Short-Term Memory Architecture for Human Re-identification. In *ECCV*.
[27] Jin Wang, Nong Sang, Zheng Wang, and Changxin Gao. 2016. Similarity learning with top-heavy ranking loss for person re-identification. *IEEE Signal Proc. Let.* 23, 1 (2016), 84–88.
[28] Jin Wang, Zheng Wang, Changxin Gao, Nong Sang, and Rui Huang. 2017. DeepList: Learning Deep Features With Adaptive Listwise Constraint for Person Reidentification. *IEEE Trans. Circuits Syst. Video Technol.* 27, 3 (2017), 513–524.
[29] Xiaogang Wang, Gianfranco Doretto, Thomas Sebastian, Jens Rittscher, and Peter Tu. 2007. Shape and appearance context modeling. In *ICCV*.
[30] Zheng Wang, Ruimin Hu, Chao Liang, Qingming Leng, and Kaimin Sun. 2014. Region-based interactive ranking optimization for person re-identification. In *PCM*.
[31] Zheng Wang, Ruimin Hu, Chao Liang, and Yi Yu. 2016. Zero-Shot Person Re-identification via Cross-View Consistency. *IEEE Trans. Multimedia* 18, 2 (2016), 260 – 272.
[32] Zheng Wang, Ruimin Hu, Yi Yu, Chao Liang, and Wenxin Huang. 2015. Multi-Level Fusion for Person Re-identification with Incomplete Marks. In *ACM MM*.
[33] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan. 2015. A Light CNN for Deep Face Representation with Noisy Labels. In *arXiv*.
[34] Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and Stan Z. Li. 2014. Salient Color Names for Person Re-identification. In *ECCV*.
[35] Mang Ye, Chao Liang, Yi Yu, Zheng Wang, Qingming Leng, Chunxia Xiao, Jun Chen, and Ruimin Hu. 2016. Person Re-identification via Ranking Aggregation of Similarity Pulling and Dissimilarity Pushing. *IEEE Trans. Multimedia* 18, 12 (2016), 2553 – 2566.
[36] Mang Ye, Andy J Ma, Liang Zheng, Jiawei Li, and Pong C Yuen. 2017. Dynamic Label Graph Matching for Unsupervised Video Re-Identification. In *ICCV*.
[37] Dong Yi, Zhen Lei, Shengcai Liao, and S. Z Li. 2014. Deep Metric Learning for Person Re-identification. In *ICPR*.
[38] Yiming Ying and Peng Li. 2012. Distance Metric Learning with Eigenvalue Optimization. *J. Mach. Learn. Res.* (2012).
[39] Li Zhang, Tao Xiang, and Shaogang Gong. 2016. Learning a Discriminative Null Space for Person Re-identification. In *CVPR*.
[40] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. 2013. Unsupervised Salience Learning for Person Re-identification. In *CVPR*.
[41] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. 2015. Scalable Person Re-identification: A Benchmark. In *ICCV*.
[42] Liang Zheng, Yi Yang, and Alexander G. Hauptmann. 2016. Person Re-identification: Past, Present and Future. In *arXiv*.
[43] Liang Zheng, Yi Yang, and Qi Tian. 2017. SIFT meets CNN: A decade survey of instance retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* (2017).
[44] Wei Shi Zheng, Shaogang Gong, and Tao Xiang. 2011. Person re-identification by probabilistic relative distance comparison. In *CVPR*.
[45] Xiangyu Zhu, Zhen Lei, Junjie Yan, and Dong Yi. 2015. High-fidelity Pose and Expression Normalization for face recognition in the wild. In *CVPR*.