Large Margin Multi-Metric Learning for Face and Kinship Verification in the Wild

Junlin Hu[1], Jiwen Lu[2], Junsong Yuan[1], Yap-Peng Tan[1]

[1]School of EEE, Nanyang Technological University, Singapore
[2]Advanced Digital Sciences Center, Singapore

Abstract. Metric learning has been widely used in face and kinship verification and a number of such algorithms have been proposed over the past decade. However, most existing metric learning methods only learn one Mahalanobis distance metric from a single feature representation for each face image and cannot deal with multiple feature representations directly. In many face verification applications, we have access to extract multiple features for each face image to extract more complementary information, and it is desirable to learn distance metrics from these multiple features so that more discriminative information can be exploited than those learned from individual features. To achieve this, we propose a new large margin multi-metric learning (LM^{3}L) method for face and kinship verification in the wild. Our method jointly learns multiple distance metrics under which the correlations of different feature representations of each sample are maximized, and the distance of each positive is less than a low threshold and that of each negative pair is greater than a high threshold, simultaneously. Experimental results show that our method can achieve competitive results compared with the state-of-the-art methods.

1 Introduction

Metric learning techniques have been widely used in many visual analysis applications such as face recognition [5, 9, 21], image classification [28], human activity recognition [27], and kinship verification [17]. Over the past decade, a large number of metric learning algorithms have been proposed and some of them have been successfully applied to face and kinship verification [5, 9, 17, 21]. In face image analysis, we usually have access to multiple feature representations for each face image and it is desirable to learn distance metrics from these multiple feature representations such that more discriminative information can be exploited than those learned from individual features. A possible solution is to concatenate different features together as a new feature vector and then apply existing metric learning algorithms directly on the concatenated vector. However, this concatenation is not physically meaningful because each feature has its own statistical characteristic, and such a concatenation ignores the diversity of multiple features and cannot effectively explore their complementary information.

In this paper, we propose a new large margin multi-metric learning (LM^{3}L) method for face and kinship verification in the wild. Instead of learning a distance

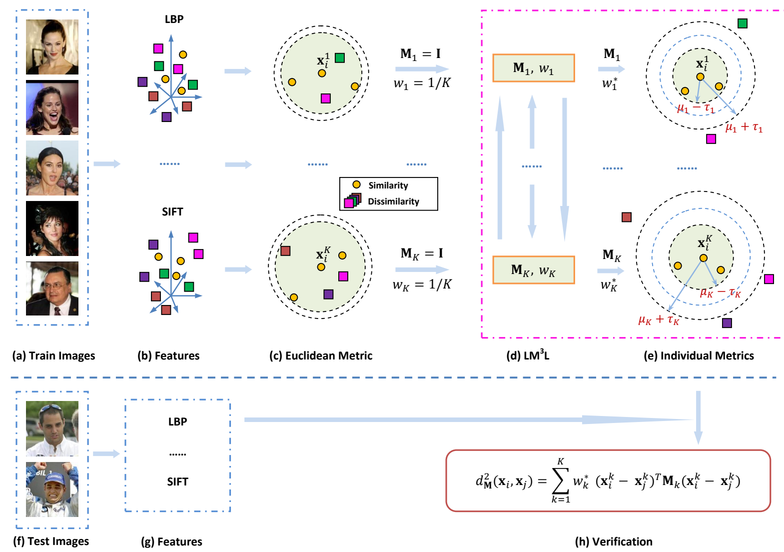

Fig. 1. Illustration of our large margin multi-metric learning method for face verification, which jointly learns multiple distance metrics, one for each feature descriptor, and collaboratively optimizes the objective function over different features. (a) A training face image set; (b) The extracted K different feature sets; (c) The distribution of these multiple feature representations in the Euclidean metric space; (d) Our LM^{3}L learning procedure; (e) The learned multiple distance metrics; (f) The test face image pair; (g) The extracted multiple feature descriptors of the test face pairs; (h) The overall distance by fusing the multiple distance metrics learned by our method.

metric with concatenated feature vectors, we jointly learn multiple distance metrics from multiple feature representations, where one metric is learned for each feature and the correlations of different feature representations of each sample are maximized, and the distance of each positive face pair is less than a smaller threshold and that of each negative pair is higher than a larger threshold, respectively. Experimental results on three widely used face datasets show that our method can obtain competitive results compared with the state-of-the-art methods. Fig. 1 illustrates the working flow of our method.

2 Related Work

Face and Kinship Verification in the Wild: In recent years, many approaches have been proposed for face and kinship verification in the wild, and they can be mainly classified into two categories: feature-based [7, 10, 37, 38] and model-based [17, 18, 33, 34]. Feature-based methods represent each face image by using a hand-crafted or learned descriptor. State-of-the-art descriptors include Gabor feature, local binary pattern (LBP) [1], locally adaptive regression kernel (LARK) [23], probabilistic elastic matching (PEM) [15], fisher vector faces [25], discriminant face descriptor [14], and spatial face region descriptor (SFRD) [5]. Representative model-based methods are subspace learning, sparse representation, metric learning, multiple kernel learning, and support vector machine. In this paper, we propose a metric learning method to learn multiple distance metrics with multiple feature representations to exploit more discriminative information for face and kinship verification in the wild.

Metric Learning: A number of metric learning algorithms have been proposed in the literature, and most of them seek an appropriate distance metric to exploit discriminative information from the training samples. Representative metric learning methods include neighborhood component analysis (NCA) [8], large margin nearest neighbor (LMNN) [29], information theoretic metric learning (ITML) [6], logistic discriminant metric learning (LDML) [9], cosine similarity metric learning (CSML) [21], KISS metric embedding (KISSME) [13], pairwise constrained component analysis (PCCA) [20], neighborhood repulsed metric learning (NRML) [17], pairwise-constrained multiple metric learning (PMML) [5], and similarity metric learning (SML) [3]. While these methods have achieved encouraging performance in face and kinship verification, most of them learn a distance metric from single feature representation and cannot deal with multiple features directly. Different from these methods, we propose a multi-metric learning method by collaboratively learning multiple distance metrics, one for each feature, to better exploit more complementary information from multiple feature representations for face and kinship verification in the wild.

3 Proposed Method

Before detailing our method, we first list the notations used in this paper. Bold capital letters, e.g., $\mathbf{X}_1$, $\mathbf{X}_2$, represent matrices, and bold lower case letters, e.g., $\mathbf{x}_1$, $\mathbf{x}_2$, represent column vectors. Given a multi-feature data set with N training samples, i.e., $\mathbf{X} = \{\mathbf{X}_k \in \mathbb{R}^{d_k \times N}\}_{k=1}^{K}$, where $\mathbf{X}_k = [\mathbf{x}_1^k, \mathbf{x}_2^k, \cdots, \mathbf{x}_N^k]$ is the feature matrix extracted from the kth feature descriptor; $\mathbf{x}_i^k$ is the feature vector of the sample $\mathbf{x}_i$ in the kth feature space, $k = 1, 2, \cdots, K$; K is the total number of features; and d_k is feature dimension of $\mathbf{x}_i^k$.

3.1 Problem Formulation

Let $\mathbf{X}_k = [\mathbf{x}_1^k, \mathbf{x}_2^k, \cdots, \mathbf{x}_N^k]$ be a feature set from the kth feature representation, the squared Mahalanobis distance between a pair of samples $\mathbf{x}_i^k$ and $\mathbf{x}_j^k$ can be computed as:

$$d^2_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k) = (\mathbf{x}_i^k - \mathbf{x}_j^k)^T \mathbf{M}_k (\mathbf{x}_i^k - \mathbf{x}_j^k), \tag{1}$$

where $\mathbf{M}_k \in \mathbb{R}^{d_k \times d_k}$ is a positive definite matrix.

We seek a distance metric $\mathbf{M}_k$ such that the squared distance $d^2_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k)$ for a face pair $\mathbf{x}_i^k$ and $\mathbf{x}_j^k$ in the kth feature space should be smaller than a given threshold $\mu_k - \tau_k$ ($\mu_k > \tau_k > 0$) if two samples are from the same subject, and larger than a threshold $\mu_k + \tau_k$ if these two samples are from different subjects, which can be formulated as the following constraints:

$$y_{ij}\big(\mu_k - d^2_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k)\big) > \tau_k, \tag{2}$$

where $y_{ij} = 1$ if $\mathbf{x}_i^k$ and $\mathbf{x}_j^k$ are from the same person, otherwise $y_{ij} = -1$.

To learn $\mathbf{M}_k$, we define the constraints in Eq. (2) by a hinge loss function, and formulate the following objective function to learn the kth distance metric:

$$\min_{\mathbf{M}_k} I_k = \sum_{i,j} h\Big(\tau_k - y_{ij}\big(\mu_k - d^2_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k)\big)\Big), \tag{3}$$

where $h(x) = \max(x, 0)$ represents the hinge loss function.

Then, our large margin multi-metric learning (LM^{3}L) method aims to learn K distance metrics $\{\mathbf{M}_k \in \mathbb{R}^{d_k \times d_k}\}_{k=1}^K$ for a multi-feature dataset, such that

1. The discriminative information from each single feature can be exploited as much as possible;
2. The differences of different feature representations of each sample in the learned distance metrics are minimized, because different features of each sample share the same semantic label.

Since the difference computation of the sample $\mathbf{x}_i$ from the kth and ℓth ($1 \le k, \ell \le K$, $k \neq \ell$) feature representations relies on the distance metrics $\mathbf{M}_k$ and $\mathbf{M}_\ell$, which could be different in dimensions, it is infeasible to compute them directly. To address this, we use an alternative constrain to reflect the relationships of different feature representations. Since the difference of $\mathbf{x}_i^k$ and $\mathbf{x}_i^\ell$, and that of $\mathbf{x}_j^k$ and $\mathbf{x}_j^\ell$ are expected to be minimized as much as possible, the distance between $\mathbf{x}_i^k$ and $\mathbf{x}_j^k$, and that of $\mathbf{x}_i^\ell$ and $\mathbf{x}_j^\ell$ are also expected to be as small as possible. Hence, we formulate the following objective function to constrain the interactions of different distance metrics in our LM^{3}L method:

$$\begin{aligned} \min_{\mathbf{M}_1,\cdots,\mathbf{M}_K} J &= \sum_{k=1}^K w_k\ I_k + \lambda \sum_{k,\ell=1,k<\ell}^K \sum_{i,j} \Big(d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k) - d_{\mathbf{M}_\ell}(\mathbf{x}_i^\ell, \mathbf{x}_j^\ell)\Big)^2, \\ \text{s.t.} \quad & \sum_{k=1}^K w_k = 1,\ w_k \ge 0,\ \lambda > 0, \end{aligned} \tag{4}$$

where w_k is a nonnegative weighting parameter to reflect the importance of the kth feature in the whole objective function, and λ weights the pairwise difference of the distance between two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ in the learned distance metrics $\mathbf{M}_k$ and $\mathbf{M}_\ell$. The physical meaning of Eq. (4) is that we aim to learn K distance metrics $\{\mathbf{M}_k\}_{k=1}^K$ under which the difference of feature representations of each pair of face samples is enforced to be as small as possible, which is consistent to the canonical correlation analysis-based multiple feature fusion approach [24].

Having obtained multiple distance metrics $\{\mathbf{M}_k\}_{k=1}^K$, the distance between two multi-feature data $\mathbf{x}_i$ and $\mathbf{x}_j$ can be computed as

$$d^2_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{K} w_k (\mathbf{x}_i^k - \mathbf{x}_j^k)^T \mathbf{M}_k (\mathbf{x}_i^k - \mathbf{x}_j^k). \tag{5}$$

The trivial solution of Eq. (4) is $w_k = 1$, which corresponds to the minimum I_k over different feature representations, and $w_k = 0$ otherwise. This solution means that only one single feature that yields the best verification accuracy is selected, which does not satisfy our objective on exploring the complementary property of multi-feature data.

To address this shortcoming, we modify w_k to be w_k^p ($p > 1$), then the new objective function is rewritten as:

$$\min_{\mathbf{M}_1,\cdots,\mathbf{M}_K} J = \sum_{k=1}^{K} w_k^p \ I_k + \lambda \sum_{k,\ell=1,k<\ell}^{K} \sum_{i,j} \Big(d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k) - d_{\mathbf{M}_\ell}(\mathbf{x}_i^\ell, \mathbf{x}_j^\ell) \Big)^2,$$
$$\text{s.t.} \quad \sum_{k=1}^{K} w_k = 1, \ w_k \geq 0, \ \lambda > 0. \tag{6}$$

3.2 Alternating Optimization

To our best knowledge, it is non-trivial to seek a global optimal solution to Eq. (6) because there are K metrics to be learned simultaneously. In this work, we employ an iterative method by using the alternating optimization method to obtain a local optimal solution. The alternating optimization learns $\mathbf{M}_k$ and w_k in an iterative manner. In our experiments, we randomly select the order of different features to start the optimization procedure and our tests show that the influence of this order is not critical to the final verification performance.

Fix $\mathbf{w} = [w_1, w_2, \cdots, w_K]$, update $\mathbf{M}_k$. With the fixed $\mathbf{w}$, we can cyclically optimize Eq. (6) over different features. We sequentially optimize $\mathbf{M}_k$ with the fixed $\mathbf{M}_1$, $\cdots$, $\mathbf{M}_{k-1}$, $\mathbf{M}_{k+1}$, $\cdots$, $\mathbf{M}_K$. Hence, Eq. (6) can be rewritten as:

$$\min_{\mathbf{M}_k} \ J = w_k^p \ I_k + \lambda \sum_{\ell=1,\ell\neq k}^{K} \sum_{i,j} \Big(d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k) - d_{\mathbf{M}_\ell}(\mathbf{x}_i^\ell, \mathbf{x}_j^\ell) \Big)^2 + A_k, \tag{7}$$

where A_k is a constant term.

To learn metric $\mathbf{M}_k$, we employ a gradient-based scheme. After some algebraic simplification, we can obtain the gradient as:

$$\frac{\partial J}{\partial \mathbf{M}_k} = w_k^p \sum_{i,j} y_{ij} h'(z) \mathbf{C}_{ij}^k \ + \ \lambda \sum_{\ell=1,\ell\neq k}^{K} \sum_{i,j} \left(1 - \frac{d_{\mathbf{M}_\ell}(\mathbf{x}_i^\ell, \mathbf{x}_j^\ell)}{d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k)} \right) \mathbf{C}_{ij}^k, \tag{8}$$

where $z = \tau_k - y_{ij}\big(\mu_k - d^2_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k)\big)$ and $\mathbf{C}_{ij}^k = (\mathbf{x}_i^k - \mathbf{x}_j^k)(\mathbf{x}_i^k - \mathbf{x}_j^k)^T$. The $\mathbf{C}_{ij}^k$ denotes the outer product of pairwise differences. $h'(x)$ is the derivative of $h(x)$, and we handle the non-differentiability of $h(x)$ at $x = 0$ by adopting a smooth hinge function as in [22, 26]. In addition, we use some derivations given as:

$$\frac{\partial}{\partial \mathbf{M}_k} d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k) = \frac{1}{2\; d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k)}\; \mathbf{C}_{ij}^k, \tag{9}$$

$$\begin{aligned} &\frac{\partial}{\partial \mathbf{M}_k}\Big(d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k) - d_{\mathbf{M}_\ell}(\mathbf{x}_i^\ell, \mathbf{x}_j^\ell)\Big)^2 \\ &\quad = 2\;\Big(d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k) - d_{\mathbf{M}_\ell}(\mathbf{x}_i^\ell, \mathbf{x}_j^\ell)\Big)\; \frac{\partial}{\partial \mathbf{M}_k} d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k) \\ &\quad = \left(1 - \frac{d_{\mathbf{M}_\ell}(\mathbf{x}_i^\ell, \mathbf{x}_j^\ell)}{d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k)}\right)\; \mathbf{C}_{ij}^k. \end{aligned} \tag{10}$$

Then, matrix $\mathbf{M}_k$ can be obtained by using a gradient descent algorithm:

$$\mathbf{M}_k = \mathbf{M}_k - \beta\; \frac{\partial J}{\partial \mathbf{M}_k}, \tag{11}$$

where β is the learning rate.

In practice, directly optimizing the Mahalanobis distance metric $\mathbf{M}_k$ may suffer slow convergence and overfitting problems if data is very high-dimensional and the number of training samples is insufficient. Therefore, we propose an alternative method to jointly perform dimensionality reduction and metric learning, which means a low-rank linear projection matrix $\mathbf{L}_k \in \mathbb{R}^{s_k \times d_k}$ $(s_k < d_k)$ is learned to project each sample $\mathbf{x}_i^k$ from the high-dimensional input space to a low-dimensional embedding space, where the distance metric $\mathbf{M}_k = {\mathbf{L}_k}^T \mathbf{L}_k$. Then, we differentiate the objective function J with respect to $\mathbf{L}_k$, and obtain the gradient as follows:

$$\frac{\partial J}{\partial \mathbf{L}_k} = 2\mathbf{L}_k \Bigg[w_k^p \sum_{i,j} y_{ij} h'(z) \mathbf{C}_{ij}^k + \lambda \sum_{\ell=1, \ell \neq k}^{K} \sum_{i,j} \left(1 - \frac{d_{\mathbf{M}_\ell}(\mathbf{x}_i^\ell, \mathbf{x}_j^\ell)}{d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k)}\right) \mathbf{C}_{ij}^k\Bigg]. \tag{12}$$

Lastly, the matrix $\mathbf{L}_k$ can be obtained by using a gradient descent rule:

$$\mathbf{L}_k = \mathbf{L}_k - \beta\; \frac{\partial J}{\partial \mathbf{L}_k}. \tag{13}$$

Fix $\mathbf{M}_k$, $k = 1, 2, \cdots, K$, update $\mathbf{w}$. Now, we update $\mathbf{w}$ with the fixed $\{\mathbf{M}_k\}_{k=1}^K$. We construct a Lagrange function as follows:

$$\begin{aligned} La(\mathbf{w}, \eta) = \sum_{k=1}^{K} w_k^p\; I_k + \lambda \sum_{k,\ell=1, k<\ell}^{K} \sum_{i,j} \Big(d_{\mathbf{M}_k}(\mathbf{x}_i^k, \mathbf{x}_j^k) - d_{\mathbf{M}_\ell}(\mathbf{x}_i^\ell, \mathbf{x}_j^\ell)\Big)^2 \\ - \eta\; \Big(\sum_{k=1}^{K} w_k - 1\Big). \end{aligned} \tag{14}$$

Algorithm 1: LM^{3}L

Input: Training set $\{\mathbf{X}_k\}_{k=1}^K$ from K views; Learning rate β; Parameter p, λ, μ_k and τ_k; Total iterative number T; Convergence error ε.
Output: Multiple metrics: $\mathbf{M}_1, \mathbf{M}_2, \cdots, \mathbf{M}_K$; and weights: $w_1, w_2, \cdots, w_K$.
Step 1 (Initialization):
 Initialize $\mathbf{L}_k = \mathbf{E}^{s_k \times d_k}$,
 $w_k = 1/K$, $k = 1, \cdots, K$.
Step 2 (Alternating optimization):
 for $t = 1, 2, \cdots, T$, **do**
  for $k = 1, 2, \cdots, K$, **do**
   Compute $\mathbf{L}_k$ by Eqs. (12) and (13).
  end for
  Compute $\mathbf{w}$ according to Eq. (17).
  Computer $J^{(t)}$ via Eq. (6).
  If $t > 1$ and $|J^{(t)} - J^{(t-1)}| < \varepsilon$
   Go to **Step 3**.
  end if
 end for
Step 3 (Output distance metrics and weights):
 $\mathbf{M}_k = {\mathbf{L}_k}^T \mathbf{L}_k$, $k = 1, 2, \cdots, K$.
 Output $\mathbf{M}_1, \mathbf{M}_2, \cdots, \mathbf{M}_K$ and $\mathbf{w}$.

Let $\frac{\partial La(\mathbf{w},\eta)}{\partial w_k} = 0$ and $\frac{\partial La(\mathbf{w},\eta)}{\partial \eta} = 0$, we have

$$\frac{\partial La(\mathbf{w}, \eta)}{\partial w_k} = p\ w_k^{p-1}\ I_k - \eta = 0, \tag{15}$$

$$\frac{\partial La(\mathbf{w}, \eta)}{\partial \eta} = \sum_{k=1}^{K} w_k - 1 = 0. \tag{16}$$

According to Eqs. (15) and (16), w_k can be updated as follows:

$$w_k = \frac{\left(1/I_k\right)^{1/(p-1)}}{\sum\limits_{k=1}^{K} \left(1/I_k\right)^{1/(p-1)}}. \tag{17}$$

We repeat the above two steps until the algorithm meets a certain convergence condition. The proposed LM^{3}L algorithm is summarized in **Algorithm 1**, where $\mathbf{E} \in \mathbb{R}^{s_k \times d_k}$ is a matrix with 1's on the diagonal and zeros elsewhere.

4 Experiments

To evaluate the effectiveness of our LM^{3}L method, we conduct face and kinship verification in the wild experiments on three real-world face datasets including the Labeled Faces in the Wild (LFW) [12], the YouTube Faces (YTF) [30],

Fig. 2. Some sample positive pairs from the LFW, YTF and KinFaceW-II datasets.

and the KinFaceW-II [17]. Fig. 2 shows some sample images from these three datasets. The parameters p, β, λ, μ_k and τ_k of our LM^3L method were empirically set as 2, 0.001, 0.1, 5 and 1 for all $k = 1, 2, \cdots, K$, respectively. The following details the experiments and results.

4.1 Datasets and Settings

LFW. The LFW dataset [12] 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. There are two training paradigms for supervised learning on this dataset: *image-restricted* and *unrestricted*. In our experiments, we use the *image-restricted* setting where only the pairwise label information is required to train our method. We follow the standard evaluation protocol on the "View 2" dataset [12] which includes 3000 matched pairs and 3000 mismatched pairs and is divided into 10 folds with each fold consisting of 300 matched (positive) pairs and 300 mismatched (negative) pairs. We use the aligned LFW-a dataset[1] for our evaluation, and crop each image into 80×150 to remove the background information. For each face image, we extracted three different features: 1) Dense SIFT (DSIFT) [16]: We densely sample SIFT descriptors on each 16×16 patch without overlapping and obtain 45 SIFT descriptors. Then, we concatenate these SIFT descriptors to form one $5,760$-dimensional feature vector; 2) LBP [1]: We divide each image into 8×15 non-overlapping blocks, where the size of each block is 10×10. Then, we extract a 59-dimensional uniform pattern LBP feature for each block and concatenate them to form a 7080-dimensional feature vector; 3) Sparse SIFT (SSIFT): We use the SSIFT feature provided by [9], which first localizes nine fixed landmarks in each image and extracts SIFT features over three scales at these landmarks, then concatenates these SIFT descriptors to form one 3456-dimensional feature vector. For these three features, we performed whitened PCA (WPCA) to project each feature into a 200 dimensional feature subspace, respectively.

[1] Available: http://www.openu.ac.il/home/hassner/data/lfwa/.

YTF. The YTF dataset [30] contains 3425 videos of 1595 different people collected from YouTube site. There are large variations in pose, illumination, and expression in each video, and the average length of each video clip is 181.3 frames. In our experiments, we follow the standard evaluation protocol and test our method for unconstrained face verification with 5000 video pairs. These pairs are equally divided into 10 folds with each fold has 250 intra-personal pairs and 250 inter-personal pairs. Similar to LFW, we also adopt the *image restricted* protocol to evaluate our method. For this dataset, we directly use three feature descriptors including LBP, Center-Symmetric LBP (CSLBP) [30] and Four-Patch LBP (FPLBP) [31] which are provided by [30]. Since all face images have been aligned by the detected facial key points, we average all the feature vectors within one video clip to form a mean feature vector. Lastly, we also use WPCA to map each feature into a 200-dimensional feature vector.

KinFaceW-II. The KinFaceW-II [17] is a kinship face dataset collected from the public figures or celebrities and their parents or children. There are four kinship relations in the KinFaceW-II datasets: Father-Son (F-S), Father-Daughter (F-D), Mother-Son (M-S) and Mother-Daughter (M-D), and each relation contains 250 pairs of kinship images. Following the experimental settings in [17], we construct 250 positive pairs (with kinship) and 250 negative pairs (without kinship) for each relation. For each face image, we also extract four types of features: LEarning-based descriptor (LE) [4], LBP, TPLBP and SIFT, and their dimensions are 200, 256, 256 and 200, respectively. We adopted the 5-fold cross validation strategy for each of the four subsets in this dataset and the finial results are reported by the mean verification accuracy.

4.2 Experimental Results on LFW

Comparison with Different Metric Learning Strategies: We first compare our method with three other different metric learning strategies: 1) Single Metric Learning (SML): we learn a single distance metric by using Eq. (3) with each feature representation; 2) Concatenated Metric Learning (CML): we first concatenate different features into a longer feature vector and then apply Eq. (3) to learn a distance metric; 3) Individual Metric Learning (IML): we learn the distance metric for each feature representation by using Eq. (3) and then use the equal weight to compute the similarity of two face images with Eq. (5). Table 1 records the verification rates with standard error of different metric learning strategies on the LFW dataset under the image restricted setting. We can see that our LM^{3}L consistently outperforms the other compared metric learning strategies in terms of the mean verification rate. Compared to SML, our LM^{3}L learns multiple distance metrics with multi-feature representations, such that more discriminative information can be exploited for verification. Compared with CML and IML, our LM^{3}L can jointly learn multiple distance metrics such that the distance metrics learned for different features can interact each other such that more complementary information can be extracted for verification.

Table 1. Comparisons of the mean verification rate (%) with different metric learning strategies on the LFW under image-restricted setting with label-free outside data.

Method	Feature	Accuracy (%)
SML	DSIFT	84.30 ± 2.17
SML	LBP	83.83 ± 1.31
SML	SSIFT	84.58 ± 1.14
CML	All	82.40 ± 1.62
IML	All	87.78 ± 1.83
LM^3L	All	$\mathbf{89.57 \pm 1.53}$

Table 2. Comparisons of the mean verification rate (%) with the state-of-the-art results on the LFW under image-restricted setting with label-free outside data, where NoF denotes the number of feature used in each method.

Method	NoF	Accuracy (%)
PCCA [20]	1	83.80 ± 0.40
PAF [35]	1	87.77 ± 0.51
CSML+SVM [21]	6	88.00 ± 0.37
SFRD+PMML [5]	8	89.35 ± 0.50
Sub-SML [3]	6	89.73 ± 0.38
DDML [11]	6	90.68 ± 1.41
VMRS [2]	10	91.10 ± 0.59
LM^3L	3	$\mathbf{89.57 \pm 1.53}$

Comparison with the State-of-the-Art Methods: We compare our LM^3L method with the state-of-the-art methods on the LFW dataset[2]. These compared methods can be categorized into metric learning based methods such as LDML [9], PCCA [20], CSML+SVM [21], DML-eig combined [36], SFRD+PMML [5], Sub-SML [3], and discriminative deep metric learning (DDML) [11]; and descriptor based methods such as Multiple LE+comp [4], Pose Adaptive Filter (PAF) [35], and high dimensional vector multiplication (VMRS) [2]. Table 2 tabulates the mean verification rate with standard error and Fig. 3 shows the ROC curves of different methods on this dataset, respectively. We can see that our LM^3L achieves competitive results with these state-of-the-art methods except VMRS [2] and DDML [11], where they run on the 10 and 6 kinds of feature, respectively.

Comparison with the Latest Multiple Metric Learning Method: We compare our LM^3L method with the latest multiple metric learning method called PMML [5]. The standard implementation of PMML was provided by the original authors. Table 3 tabulates the mean verification rate with standard error on this dataset. We can clearly see that our LM^3L significantly outperforms PMML on the LFW dataset. This is because our LM^3L can adaptively learn different weights to reflect the different importance of different features while

[2] Available: http://vis-www.cs.umass.edu/lfw/results.html.

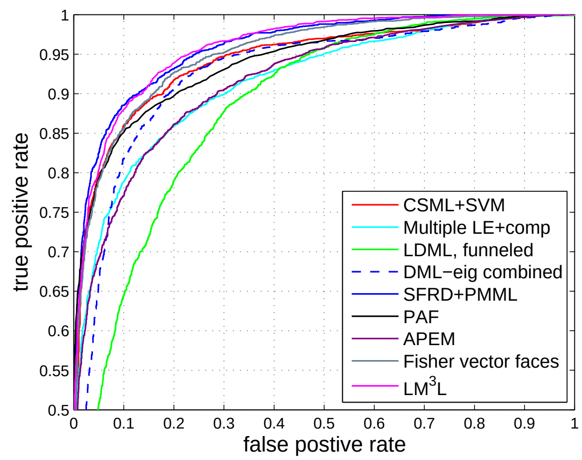

Fig. 3. Comparisons of ROC curves between our LM^{3}L and the state-of-the-art methods on the LFW under image-restricted setting with label-free outside data.

Table 3. Comparison with the latest multiple metric learning method on the LFW under image-restricted setting with label-free outside data.

Method	Accuracy (%)
PMML [5]	85.23 ± 1.69
LM^{3}L	$\mathbf{89.57 \pm 1.53}$

PMML assigns equal weights to different features, such that our method can better exploit the complementary information.

4.3 Experimental Results on YTF

Comparison with Different Metric Learning Strategies: Similar to LFW, we also compare our method with different metric learning strategies such as SML, CML, and IML on the YTF dataset. Table 4 records the verification rates of different metric learning strategies on the YTF dataset under the image restricted setting. We can also see that our LM^{3}L consistently outperforms the other metric learning strategies in terms of the mean verification rate.

Comparison with the State-of-the-Art Methods: We compare our method with the state-of-the-art methods on the YTF dataset. These compared methods include Matched Background Similarity (MBGS) [30], APEM [15], STFRD+PMML [5], MBGS+SVM⊖ [32], VSOF+OSS (Adaboost) [19], and D-DML [11]. Table 5 records the mean verification rate with the standard error, and Fig. 4 shows the ROC curves of our LM^{3}L and the state-of-the-art methods on the YTF dataset, respectively. We can observe that our LM^{3}L method achieves competitive result compared with these state-of-the-art methods on this dataset under the image restricted setting.

Table 4. Comparison of the mean verification rate with standard error (%) with different metric learning strategies on the YTF under the image restricted setting.

Method	Feature	Accuracy (%)
SML	CSLBP	73.66 ± 1.52
SML	FPLBP	75.02 ± 1.67
SML	LBP	78.46 ± 0.94
CML	All	75.36 ± 2.37
IML	All	80.12 ± 1.33
LM^{3}L	All	$\mathbf{81.28 \pm 1.17}$

Table 5. Comparisons of the mean verification rate with standard error (%) with the state-of-the-art results on the YTF under the image restricted setting.

Method	Accuracy (%)
MBGS (LBP) [30]	76.40 ± 1.80
APEM (LBP) [15]	77.44 ± 1.46
APEM (fusion) [15]	79.06 ± 1.51
STFRD+PMML [5]	79.48 ± 2.52
MBGS+SVM$\ominus$ [32]	79.48 ± 2.52
VSOF+OSS (Adaboost) [19]	79.70 ± 1.80
DDML (combined) [11]	82.34 ± 1.47
LM^{3}L	$\mathbf{81.28 \pm 1.17}$

Comparison with the Latest Multiple Metric Learning Method: Table 6 shows the mean verification rate with standard error of our proposed method and PMML method on the YTF dataset. We can clearly see that our LM^{3}L outperforms PMML on this dataset.

4.4 Experimental Results on KinFaceW-II

Comparison with Different Metric Learning Strategies: We first compare our method with SML, CML, and IML on the KinFaceW-II dataset. Table 7 records the mean verification rates of different metric learning strategies on the KinFaceW-II dataset for four relations, respectively. We can also see that our LM^{3}L consistently outperforms the other compared metric learning strategies in terms of the mean verification rate.

Comparison with the State-of-the-Art Methods: We compare our method with the state-of-the-art methods on the KinFaceW-II dataset. These

Table 6. Comparison with the existing multiple metric learning method on the YTF under the image restricted setting.

Method	Accuracy (%)
PMML [5]	76.60 ± 1.62
LM^{3}L	$\mathbf{81.28 \pm 1.17}$

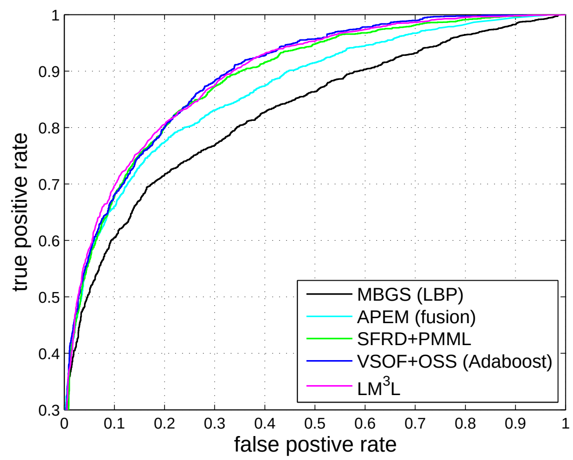

Fig. 4. Comparisons of ROC curves between our LM^{3}L and the state-of-the-art methods on the YTF under the image restricted setting.

Table 7. Comparisons of the mean verification rate (%) with different metric learning strategies on the KinFaceW-II dataset.

Method	Feature	F-S	F-D	M-S	M-D	Mean
SML	LE	76.2	70.1	72.4	71.8	72.6
SML	LBP	66.9	65.5	63.1	68.3	66.0
SML	TPLBP	71.8	63.3	63.0	67.6	66.4
SML	SIFT	68.1	63.8	67.0	63.9	65.7
CML	All	76.3	67.5	74.3	75.4	73.4
IML	All	79.4	71.5	76.3	77.3	76.1
LM^{3}L	All	**82.4**	**74.2**	**79.6**	**78.7**	**78.7**

Table 8. Comparisons of the mean verification rate (%) with the state-of-the-art methods on the KinFaceW-II dataset.

Method	Feature	F-S	F-D	M-S	M-D	Mean
PMML [5]	All	77.7	72.4	76.3	74.8	75.3
MNRML [17]	All	76.9	74.3	77.4	77.6	76.5
LM^{3}L	All	**82.4**	**74.2**	**79.6**	**78.7**	**78.7**

compared methods include MNRML [17] and PMML [5]. Table 8 reports the mean verification rates of our method and these methods. We can observe that our LM^{3}L achieves about 2.0% improvement over the current state-of-the-art result on this dataset for kinship verification.

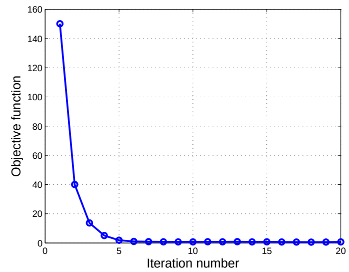

Fig. 5. The value of objective function of LM^{3}L versus different number of iterations on the LFW dataset.

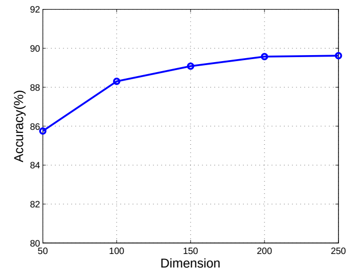

Fig. 6. The mean verification rate of LM^{3}L versus different feature dimensions on the LFW dataset.

4.5 Parameter Analysis

Since LM^{3}L is an iterative algorithm, we first evaluate its convergence with different number of iterations. Fig. 5 shows the value of the objective function of LM^{3}L versus different number of iterations on the LFW dataset. We can see that the convergence speed of LM^{3}L is fast and it converges in $5-6$ iterations.

Lastly, we evaluate the performance of LM^{3}L versus different feature dimensions. Fig. 6 shows the mean verification rate versus different feature dimensions on the LFW dataset. We can see that the proposed LM^{3}L method can achieve stable performance when the feature dimension reaches 150.

5 Conclusion and Future Work

In this paper, we have proposed a large margin multi-metric learning (LM^{3}L) method for face and kinship verification. Our method has jointly learned multiple distance metrics under which more discriminative and complementary information can be exploited. Experimental results show that our method can achieve competitive results compared with the state-of-the-art methods. For future work, we are interested to apply our method to other computer vision applications such as visual tracking and action recognition to further show its effectiveness.

Acknowledgement. This work was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by a grant from the Singapore National Research Foundation. This grant is administered by the Interactive & Digital Media Programme Office at the Media Development Authority, Singapore.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. TPAMI **28** (2006) 2037–2041

2. Barkan, O., Weill, J., Wolf, L., Aronowitz, H.: Fast high dimensional vector multiplication face recognition. In: ICCV (2013) 1960–1967
3. Cao, Q., Ying, Y., Li, P.: Similarity metric learning for face recognition. In: ICCV (2013) 2408–2415
4. Cao, Z., Yin, Q., Tang, X., Sun, J.: Face recognition with learning-based descriptor. In: CVPR (2010) 2707–2714
5. Cui, Z., Li, W., Xu, D., Shan, S., Chen, X.: Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In: CVPR (2013) 3554–3561
6. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: ICML (2007) 209–216
7. Fang, R., Tang, K., Snavely, N., Chen, T.: Towards computational models of kinship verification. In: ICIP (2010) 1577–1580
8. Goldberger, J., Roweis, S.T., Hinton, G.E., Salakhutdinov, R.: Neighbourhood components analysis. In: NIPS (2004) 513–520
9. Guillaumin, M., Verbeek, J.J., Schmid, C.: Is that you? metric learning approaches for face identification. In: ICCV (2009) 498–505
10. Guo, G., Wang, X.: Kinship measurement on salient facial features. TIM **61** (2012) 2322–2325
11. Hu, J., Lu, J., Tan, Y.P.: Discriminative deep metric learning for face verification in the wild. In: CVPR (2014) 1875–1882
12. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst (2007)
13. Köstinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: CVPR (2012) 2288–2295
14. Lei, Z., Pietikainen, M., Li, S.Z.: Learning discriminant face descriptor. TPAMI **36** (2014) 289–302
15. Li, H., Hua, G., Lin, Z., Brandt, J., Yang, J.: Probabilistic elastic matching for pose variant face verification. In: CVPR (2013) 3499–3506
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. IJCV **60** (2004) 91–110
17. Lu, J., Hu, J., Zhou, X., Shang, Y., Tan, Y.P., Wang, G.: Neighborhood repulsed metric learning for kinship verification. In: CVPR (2012) 2594–2601
18. Lu, J., Zhou, X., Tan, Y.P., Shang, Y., Zhou, J.: Neighborhood repulsed metric learning for kinship verification. TPAMI **36** (2014) 331–345
19. Mendez-Vazquez, H., Martinez-Diaz, Y., Chai, Z.: Volume structured ordinal features with background similarity measure for video face recognition. In: ICB (2013) 1–6
20. Mignon, A., Jurie, F.: Pcca: A new approach for distance learning from sparse pairwise constraints. In: CVPR (2012) 2666–2672
21. Nguyen, H.V., Bai, L.: Cosine similarity metric learning for face verification. In: ACCV (2010) 709–720
22. Rennie, J.D.M., Srebro, N.: Fast maximum margin matrix factorization for collaborative prediction. In: ICML (2005) 713–719
23. Seo, H.J., Milanfar, P.: Face verification using the lark representation. TIFS **6** (2011) 1275–1286
24. Sharma, A., Kumar, A., Daume III, H., Jacobs, D.: Generalized multiview analysis: a discriminative latent space. In: CVPR (2012) 1867–1875
25. Simonyan, K., Parkhi, O.M., Vedaldi, A., Zisserman, A.: Fisher vector faces in the wild. In: BMVC (2013) 1–12

26. Torresani, L., Lee, K.C.: Large margin component analysis. In: NIPS (2006) 1385–1392
27. Tran, D., Sorokin, A.: Human activity recognition with metric learning. In: ECCV (2008) 548–561
28. Wang, Z., Hu, Y., Chia, L.T.: Image-to-class distance metric learning for image classification. In: ECCV (2010) 706–719
29. Weinberger, K.Q., Blitzer, J., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. In: NIPS (2005)
30. Wolf, L., Hassner, T., Maoz, I.: Face recognition in unconstrained videos with matched background similarity. In: CVPR (2011) 529–534
31. Wolf, L., Hassner, T., Taigman, Y.: Descriptor based methods in the wild. In: ECCVW (2008)
32. Wolf, L., Levy, N.: The svm-minus similarity score for video face recognition. In: CVPR (2013) 3523–3530
33. Xia, S., Shao, M., Fu, Y.: Kinship verification through transfer learning. In: IJCAI (2011) 2539–2544
34. Xia, S., Shao, M., Luo, J., Fu, Y.: Understanding kin relationships in a photo. TMM **14** (2012) 1046–1056
35. Yi, D., Lei, Z., Li, S.Z.: Towards pose robust face recognition. In: CVPR (2013) 3539–3545
36. Ying, Y., Li, P.: Distance metric learning with eigenvalue optimization. JMLR **13** (2012) 1–26
37. Zhou, X., Hu, J., Lu, J., Shang, Y., Guan, Y.: Kinship verification from facial images under uncontrolled conditions. In: ACM MM (2011) 953–956
38. Zhou, X., Lu, J., Hu, J., Shang, Y.: Gabor-based gradient orientation pyramid for kinship verification under uncontrolled environments. In: ACM MM (2012) 725–728