

Shrinkage Expansion Adaptive Metric Learning

Qilong Wang^{1,3}, Wangmeng Zuo², Lei Zhang³, and Peihua Li¹

¹ School of Information and Communications Engineering,
Dalian University of Technology, China

² School of Computer Science and Technology, Harbin Institute of Technology, China

³ Department of Computing, Hong Kong Polytechnic University, Hong Kong
{csqlwang, cswmzuo}@gmail.com, cslzhang@comp.polyu.edu.hk,
peihuali@dlut.edu.cn

Abstract. Conventional pairwise constrained metric learning methods usually restrict the distance between samples of a similar pair to be lower than a fixed upper bound, and the distance between samples of a dissimilar pair higher than a fixed lower bound. Such fixed bound based constraints, however, may not work well when the intra- and inter-class variations are complex. In this paper, we propose a shrinkage expansion adaptive metric learning (SEAML) method by defining a novel shrinkage-expansion rule for adaptive pairwise constraints. SEAML is very effective in learning metrics from data with complex distributions. Meanwhile, it also suggests a new rule to assess the similarity between a pair of samples based on whether their distance is shrunk or expanded after metric learning. Our extensive experimental results demonstrated that SEAML achieves better performance than state-of-the-art metric learning methods. In addition, the proposed shrinkage-expansion adaptive pairwise constraints can be readily applied to many other pairwise constrained metric learning algorithms, and boost significantly their performance in applications such as face verification on LFW and PubFig databases.

Keywords: Shrinkage-expansion rule, adaptive bound constraints, pairwise constrained metric learning, face verification.

1 Introduction

Distance metric learning aims to learn an appropriate distance metric by taking advantages of the intrinsic structure of training data. Numerous metric learning methods have been proposed for a variety of computer vision applications such as face verification [10,17,27], object classification [26], image annotation [7,33], and visual tracking [21], etc. The information conveyed by training data can be generally represented as triplet and pairwise constraints. Triplet constraints based metric learning approaches, including large margin nearest neighbor (LMNN) [35], BoostMetric [31] and FrobMetric [30], restrict that for each triplet the distance between a pair of samples from the same class should be smaller than the distance between a pair of samples from different classes. Pairwise constraints are more pervasive in metric learning. In many applications such as face verification, only pairwise constraints are available from the training data. For example, under the restricted setting of the LFW face database [15], the only known information is whether a pair of face images is matched or not.

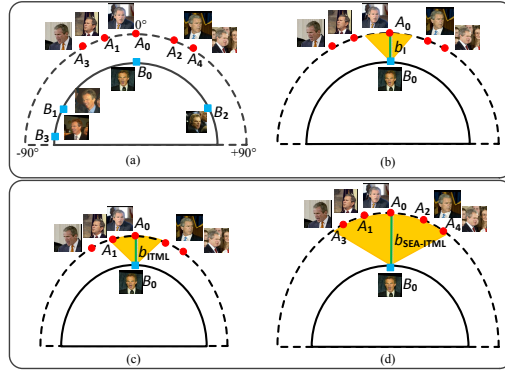


Fig. 1. An illustrative example of face verification by ITML and SEA-ITML. (a) The face image manifolds of two persons with pose variations. (b) The verification results with the decision threshold b_I before metric learning. $A_1 \sim A_4$ are all wrongly verified to be dissimilar with A_0 . (c) The verification results with the decision threshold b_{ITML} after metric learning by ITML [6] with fixed bound constraints. A_1 now is correctly verified, but $A_2 \sim A_4$ are still wrongly verified. (d) The verification results with the decision threshold $b_{SEA-ITML}$ after metric learning by ITML with the proposed shrinkage expansion adaptive constraints (SEA-ITML). $A_1 \sim A_4$ can all be correctly verified.

Most pairwise constrained metric learning methods restrict the distance between samples of a similar pair to be lower than a fixed upper bound and the distance between samples of a dissimilar pair higher than a fixed lower bound. Davis et al. [6] proposed an information theoretic metric learning (ITML) method, which restricts that the distances for all similar pairs should be smaller than an upper bound u and the distances for all dissimilar pairs should be larger than a lower bound l . Guillaumin et al. [10] learned a Mahalanobis distance metric via discriminative linear logistic regression. This method intends to find a bound b so that all distances of similar pairs can be smaller than those of dissimilar pairs. Li et al. [22] learned a second-order discriminant function instead of Mahalanobis distance for verification problem. They confined the function values of all similar pairs to be smaller than -1 and those of all dissimilar pairs larger than 1 .

The above fixed bound based constraints may fail to learn effective metrics when the intra- and inter-class variations are complex. In face recognition or verification, it is commonly accepted that the face images of a person with pose and lighting variations are in a nonlinear manifold [12]. The face manifolds of different persons are similar, while the challenge lies in that the intra-class variation of face images is often very complex. Figure 1 (a) illustrates the face manifolds of two persons, Bush ($A_0 \sim A_4$) and Blair ($B_0 \sim B_3$). Figure 1 (b) illustrates the face verification results before metric learning. To avoid false acceptance, the decision threshold can be set as b_I based on the distance between A_0 and B_0 . One can see that although A_0 and B_0 can be correctly verified as different subjects by b_I , $A_1 \sim A_4$ will be wrongly verified as different subjects from A_0 since their distances to A_0 are higher than b_I . Figure 1 (c) illustrates the verification results after metric learning by using the well-known ITML [6]. We can see that the manifolds shrink along the horizontal direction while keeping unchanged

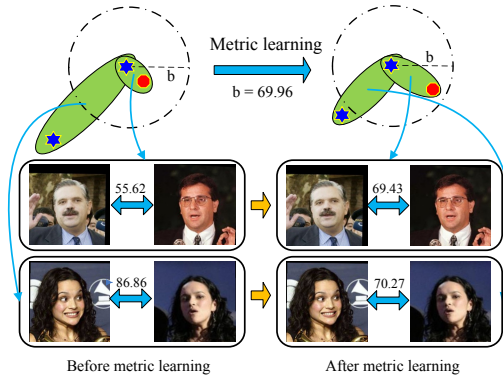


Fig. 2. A real failure case of face verification on the LFW database by conventional methods which only use a threshold b to decide if the pair is a match or non-match after metric learning. By the proposed SEAML, we can classify the pair by exploiting the shrinkage-expansion property of its sample distance, and thus both the two pairs can be correctly classified.

in the vertical direction. This time A_1 can be correctly verified by using the decision threshold b_{ITML} ; however, the other images $A_2 \sim A_4$ still cannot be correctly verified.

To tackle the above mentioned problem, in this paper we propose a self-adaptive shrinkage-expansion rule based on the original Euclidean distances of the pairs, which relaxes the fixed bound based constraints by shrinking the distances between samples of similar pairs and expanding the distances between samples of dissimilar pairs. The details can be found in Section 3.1. As illustrated in Figure 1 (d), by means of SEA-ITML (i.e., ITML with the proposed shrinkage-expansion adaptive constraints; please refer to Section 3.3 for details), $A_1 \sim A_4$ all can be correctly verified by using the decision threshold $b_{SEA-ITML}$, improving significantly the verification performance. According to our experiments, ITML [6] with the proposed pairwise constraints can boost its verification performance from 76.2% to 82.4% and from 69.3% to 77.8% on the LFW and PubFig face databases, respectively (please refer to Section 5 for details). Since ITML and SEA-ITML use the same regularization and loss terms, the benefit shall be owed to that the proposed shrinkage-expansion adaptive constraints is more effective to learn the desired metrics from data with complex distributions.

Another distinct advantage of the proposed shrinkage expansion adaptive metric learning (SEAML) method is that the shrinkage-expansion adaptive constraints suggest a novel paradigm to assess the similarity between a pair of samples. Given a pair of samples, conventional metric learning approaches usually compute their Mahalanobis distance d , and judge if they are from the same class based on whether d is lower or higher than the decision threshold b . Different from these approaches, SEAML learns a Mahalanobis distance metric to shrink the distances between samples of similar pairs and expand the distances between samples of dissimilar pairs. Thus, SEAML allows us to make the decision by considering both b and the changes between the distances before and after metric learning. Figure 2 shows two pairs of images from the LFW face database. After metric learning, the Mahalanobis distance of a similar (dissimilar)

pair is still higher (lower) than the threshold $b = 69.96$. With the conventional decision paradigm, these two pairs would be misclassified. Fortunately, one can observe that after metric learning, the distance of the similar pair shrinks from 86.86 to 70.27, while the distance of the dissimilar pair expands from 55.62 to 69.43. Thus, when the distance between samples of a pair is close to the decision threshold b , we can classify the pair according to the shrinkage-expansion property of its sample distance. In this way, both the two pairs in Figure 2 can be correctly classified by the proposed SEAML method. The details of the verification rule will be given in Section 4.

Our extensive experimental results on 12 UCI datasets [2], LFW [15] and PubFig [19] face databases show that SEAML outperforms many state-of-the-art metric learning methods. The rest of this paper is organized as follows. In Section 2, we review the literature related to our work. Section 3 presents the proposed shrinkage expansion adaptive metric learning framework and we introduce a novel verification paradigm of SEAML in Section 4. In Section 5, we show experimental results of the proposed method on both classification and face verification tasks. Finally, we conclude our paper in Section 6.

2 Related Work

Most of the existing metric learning methods learn metrics with triplet or pairwise constraints. The classical triplet constrained metric learning method LMNN [35] has shown powerful classification capability with the k nearest neighbor (kNN) classifier. Motivated by LMNN, Shen et al. [31] used the exponential loss instead of hinge loss in LMNN, resulting in a BoostMetric method. Based on LMNN and BoostMetric, Shen et al. [30] proposed the FrobMetric method, which utilizes the Frobenius norm regularizer to make metric learning more efficient and scalable. For kNN classification, the proposed SEAML shares similar philosophy to LMNN [35], BoostMetric [31] and FrobMetric [30]. For each triplet $(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$ (the class label of \mathbf{x}_i is the same as that of \mathbf{x}_j but different from that of \mathbf{x}_k), methods in [30,31,35] restrict that the learned distance $D_M(\mathbf{x}_i, \mathbf{x}_j)$ between \mathbf{x}_i and \mathbf{x}_j should be smaller than the distance $D_M(\mathbf{x}_i, \mathbf{x}_k)$ between \mathbf{x}_i and \mathbf{x}_k . From the perspective of the proposed SEAML, this triplet constraint can be regarded as that $D_M(\mathbf{x}_i, \mathbf{x}_k)$ is used as a locally adaptive upper bound of $D_M(\mathbf{x}_i, \mathbf{x}_j)$, or $D_M(\mathbf{x}_i, \mathbf{x}_j)$ is used as a locally adaptive lower bound of $D_M(\mathbf{x}_i, \mathbf{x}_k)$. However, [30,31,35] cannot be applied to tasks where only pairwise constraints are available [28]. Unlike [30,31,35], SEAML uses the Euclidean distance of a pair to adaptively set the lower and upper bounds, and thus can work in more general cases.

Some pairwise constrained metric learning methods [8,9,11] also need the label information of each sample. Neighborhood components analysis [9] learns a metric for nearest neighbor classifier (NNC) by finding a linear transformation of input data such that the average leave-one-out classification performance is maximized in the transformed space. Globerson et al. [8] proposed to learn a metric by maximizing the distances between different classes while collapsing the intra-class distances to zero. Huang et al. [11] proposed an ensemble metric learning method that restricts the intra-class distances to be smaller than inter-class distances. Similar to triplet constrained metric learning methods, they work in a fully supervised learning manner.

Although many pairwise constrained metric learning approaches, such as ITML [6], LDML [10], work in a weakly supervised manner, these methods adopt the fixed bound based constraints, making them less effective when the intra- and inter-class variations are complex. To the best of our knowledge, thus far only Li et al. [23] relaxed the fixed bound problem by learning a locally adaptive decision function to distinguish similar pairs and dissimilar ones. The proposed SEAML method in this paper is essentially different from the method in [23]. First, SEAML uses a shrinkage-expansion rule to set the adaptive pairwise constraints. Second, it suggests a novel verification paradigm by considering both the decision threshold and the shrinkage-expansion property of the distance with the learned metric. Finally, SEAML learns a Mahalanobis distance metric while [23] learns a second-order decision function. In recent years, some multiple metrics learning methods [29,34] have been proposed to handle data with complex distributions. They are distinctly different from SEAML which learns a single metric but with adaptive constraints. One may refer to [3] for more details on the recent progresses on metric learning.

3 Shrinkage Expansion Adaptive Metric Learning

In Section 3.1, we first introduce the formulation of shrinkage-expansion adaptive pairwise constraints, resulting in a shrinkage expansion adaptive metric learning (SEAML) framework. In Section 3.2, we substantiate a SEAML algorithm by using the squared Frobenius norm (F-norm) regularizer and the hinge loss, which can be efficiently optimized by alternating between SVM training [4] and projection on the cone of all positive semidefinite (PSD) matrices. Section 3.3 shows that the proposed shrinkage-expansion rule can be flexibly embedded into many existing pairwise-based metric learning methods, such as ITML [6].

3.1 Shrinkage-Expansion Adaptive Constraints

Denote the set of similar pairs by $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{x}_j) : \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to the same class}\}$, and the set of dissimilar pairs by $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{x}_j) : \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to different classes}\}$. Metric learning aims to learn a PSD matrix $\mathbf{M} \in \mathbb{R}^{d \times d}$ which characterizes the Mahalanobis distance:

$$D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j). \quad (1)$$

To learn a proper Mahalanobis distance metric, conventional pairwise constraints usually require that the learned distances of similar pairs should be lower than some fixed upper bound u , while those of dissimilar pairs should be higher than some fixed lower bound l , which can be formulated as:

$$\begin{aligned} D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) &\leq u \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}, \\ D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) &\geq l \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}. \end{aligned} \quad (2)$$

The above fixed bound based constraints may not work well for data with complex distributions, as discussed in Section 1. To address this issue, we propose to relax the fix

bound based constraints by introducing an adaptive shrinkage-expansion rule to design pairwise constraints. Given a similar/dissimilar pair $(\mathbf{x}_i, \mathbf{x}_j)$, the Euclidean distance $D_{ij} = D_{\mathbf{I}}(\mathbf{x}_i, \mathbf{x}_j)$ is used as a reference to guide the shrinkage/expansion, and thus the adaptive upper/lower bound can be defined as a function of D_{ij} , i.e., $f(D_{ij})$. One obvious principle for the design of $f(D_{ij})$ is : the larger the distance of a similar pair, the more $f(D_{ij})$ should shrink from D_{ij} , while the smaller the distance of a dissimilar pair, the more $f(D_{ij})$ should expand from D_{ij} . Based on this principle, we define the following shrinkage-expansion rule to compute the adaptive upper/lower bounds for similar/dissimilar pairs:

$$\begin{aligned} f_s(D_{ij}) &= D_{ij} - (D_{ij}^{\frac{1}{N_s}}/D_c) \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}. \\ f_d(D_{ij}) &= D_{ij} + (D_c/D_{ij}^{\frac{1}{N_d}}) \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}. \end{aligned} \quad (3)$$

where $D_c \geq 1$ is a constant. $N_s \geq 1$ and $N_d \geq 1$ are scale factors which control the level of shrinkage and expansion, respectively.

From Eq. (3), we can see that the larger the N_s , the slower $f_s(D_{ij})$ will shrink, while the larger the N_d , the faster $f_d(D_{ij})$ will expand. In this paper, we set D_c as the maximal distance (denoted by D_{max}) among all pairs used in training data and set $N_s = 1$ and $N_d = 1/\log(\frac{D_c}{D_c-2})$. N_d is abbreviated as n in the following Eq. (4). Thus, we obtain the following shrinkage-expansion adaptive constraints:

$$\begin{aligned} f_s(D_{ij}) &= D_{ij} - (D_{ij}/D_{max}) \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}. \\ f_d(D_{ij}) &= D_{ij} + (D_{max}/\sqrt[n]{D_{ij}}) \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}. \end{aligned} \quad (4)$$

We set $N_s = 1$ to ensure that $f_s(D_{ij})$ can shrink the fastest (note that $N_s < 1$ cannot guarantee $f_s(D_{ij})$ to be always positive). We set $N_d = 1/\log(\frac{D_c}{D_c-2})$ to ensure rapid expansion of $f_d(D_{ij})$, while it allows us to automatically distinguish similar and dissimilar pairs by D_c , i.e., $f_s(D_{ij}) < D_c$ and $f_d(D_{ij}) > D_c$, $\forall(i, j)$. In our experiments, it is found that when the constant $D_c > 2$, it has little influence on the classification and verification performance.

The proposed shrinkage-expansion rule can adaptively determine the pairwise constraints, and thus can effectively distinguish between similar pairs and dissimilar pairs. Based on the proposed shrinkage-expansion rule, we can formulate the SEAML framework as follows:

$$\begin{aligned} \min_{\mathbf{M}, \boldsymbol{\xi}} \quad & r_1(\mathbf{M}) + C \cdot r_2(\boldsymbol{\xi}) \\ \text{s.t.} \quad & D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \leq f_s(D_{ij}) + \xi_{ij} \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}, \\ & D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \geq f_d(D_{ij}) - \xi_{ij} \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}, \\ & \xi_{ij} \geq 0, \forall(i, j), \mathbf{M} \succcurlyeq 0. \end{aligned} \quad (5)$$

where r_1 is the regularizer on \mathbf{M} , r_2 is the loss term on $\boldsymbol{\xi}$, and ξ_{ij} is introduced as the soft penalty on the pairwise inequality constraint. Many existing regularizers (e.g., squared F-norm and LogDet divergence) and loss functions (e.g., hinge loss and logistic loss) can be used in the proposed SEAML framework.

3.2 SEAML with Squared F-Norm Regularizer

The studies in [18] have shown that the squared F-norm regularizer can achieve good performance in object recognition, and it is strictly convex. On the other hand, it is known that the hinge loss has good generalization performance and it is adopted in LMNN and SVM. Thus, we adopt the squared F-Norm regularizer and hinge loss function in the proposed SEAML framework, resulting in the following metric learning model:

$$\begin{aligned} \min_{\mathbf{M}, \xi} \quad & \frac{1}{2} \|\mathbf{M}\|_F^2 + C \sum_{ij} \xi_{ij} \\ \text{s.t.} \quad & D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \leq f_s(D_{ij}) + \xi_{ij} \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}, \\ & D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \geq f_d(D_{ij}) - \xi_{ij} \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}, \\ & \xi_{ij} \geq 0, \forall (i, j), \mathbf{M} \succcurlyeq 0. \end{aligned} \quad (6)$$

We can rewrite Eq. (6) as follows:

$$\begin{aligned} \min_{\mathbf{M}, \xi} \quad & \frac{1}{2} \|\mathbf{M}\|_F^2 + C \sum_{ij} \xi_{ij} \\ \text{s.t.} \quad & l_{ij} D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \geq f(D_{ij}, l_{ij}) - \xi_{ij}, \\ & \xi_{ij} \geq 0, \forall (i, j), \mathbf{M} \succcurlyeq 0. \end{aligned} \quad (7)$$

where $l_{ij} = -1$, $f(D_{ij}, l_{ij}) = -f_s(D_{ij})$ if $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}$ and $l_{ij} = 1$, $f(D_{ij}, l_{ij}) = f_d(D_{ij})$ if $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}$. For simplicity, hereafter $f(D_{ij}, l_{ij})$ is abbreviated as f_{ij} .

Let $\mathbf{Z}_{ij} = (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T$. We can rewrite Eq. (1) in the form of inner product: $D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) = \langle \mathbf{Z}_{ij}, \mathbf{M} \rangle = \text{tr}(\mathbf{Z}_{ij} \mathbf{M})$. Then, the Lagrange function of Eq. (7) can be expressed as follows:

$$\begin{aligned} \mathcal{L}(\mathbf{M}, \mathbf{Y}, \beta, \gamma, \xi) = \quad & \frac{1}{2} \|\mathbf{M}\|_F^2 + C \sum_{ij} \xi_{ij} - \langle \mathbf{Y}, \mathbf{M} \rangle \\ & - \sum_{ij} \beta_{ij} [l_{ij} \langle \mathbf{Z}_{ij}, \mathbf{M} \rangle - f_{ij} + \xi_{ij}] - \sum_{ij} \gamma_{ij} \xi_{ij} \\ \text{s.t.} \quad & \beta_{ij} \geq 0, \gamma_{ij} \geq 0, \xi_{ij} \geq 0 \quad \forall (i, j), \mathbf{M}, \mathbf{Y} \succcurlyeq 0. \end{aligned} \quad (8)$$

where β_{ij} , γ_{ij} , and \mathbf{Y} are the Lagrange multipliers. Based on the Karush-Kuhn-Tucker (KKT) conditions, we obtain the Lagrange dual problem of Eq. (7):

$$\begin{aligned} \max_{\mathbf{Y}, \beta} \quad & -\frac{1}{2} \left\| \sum_{ij} \beta_{ij} l_{ij} \mathbf{Z}_{ij} + \mathbf{Y} \right\|_F^2 + \sum_{ij} f_{ij} \beta_{ij} \\ \text{s.t.} \quad & 0 \leq \beta_{ij} \leq C, \forall (i, j), \mathbf{Y} \succcurlyeq 0. \end{aligned} \quad (9)$$

Please refer to the supplementary material for the detailed deduction of the Lagrange dual.

The problem in Eq. (9) involves the joint optimization of PSD matrix \mathbf{Y} and vector β , which can be solved by iterating between two procedures. First, by fixing \mathbf{Y} , the subproblem in Eq. (9) can be reformulated as a quadratic programming (QP) problem:

$$\begin{aligned} \max_{\beta} \quad & -\frac{1}{2} \sum_{ij} \sum_{pq} \beta_{ij} \beta_{pq} l_{ij} l_{pq} \langle \mathbf{Z}_{ij}, \mathbf{Z}_{pq} \rangle + \sum_{ij} f_{ij} \beta_{ij} \\ \text{s.t.} \quad & 0 \leq \beta_{ij} \leq C, \forall (i, j). \end{aligned} \quad (10)$$

where $g_{ij} = f_{ij} - l_{ij}(\langle \mathbf{Z}_{ij}, \mathbf{Y} \rangle)$. Clearly, we can use the off-the-shelf SVM solver [4] to obtain the optimal solution to the subproblem on β . After updating β , the subproblem on \mathbf{Y} can be reformulated as the projection of $\hat{\mathbf{Y}}$ on the cones of PSD matrices:

$$\begin{aligned} \min_{\mathbf{Y}} \frac{1}{2} \|\mathbf{Y} - \hat{\mathbf{Y}}\|_F^2 \\ \text{s.t. } \mathbf{Y} \succeq 0. \end{aligned} \quad (11)$$

where $\hat{\mathbf{Y}} = -\sum_{ij} \beta_{ij} l_{ij} \mathbf{Z}_{ij}$. Let $\hat{\mathbf{Y}} = \mathbf{U}\Sigma\mathbf{V}^T$ be the SVD of $\hat{\mathbf{Y}}$, where Σ is the diagonal matrix of eigenvalues, and \mathbf{U} is an orthonormal matrix consisting of the corresponding eigenvectors. The closed form solution of \mathbf{Y} can then be represented as $\mathbf{Y} = \mathbf{U}\Sigma_+\mathbf{V}^T$, where $\Sigma_+ = \max(\mathbf{0}, \Sigma)$.

We iteratively update β and \mathbf{Y} until the convergence of the objective function. Based on [5], it is guaranteed that the proposed alternating minimization approach would make the objective function in Eq. (9) converge to the global optimum. Finally, the learned matrix \mathbf{M} can be obtained by $\sum_{ij} \beta_{ij} l_{ij} \mathbf{Z}_{ij} + \mathbf{Y}$. The proposed SEAML algorithm is summarized in Algorithm 1. The time complexity of training in SEAML is $O(L(dN^2 + d^3))$, where L is the iteration number, d is the feature dimension, and N is the number of pairs.

3.3 ITML with Shrinkage-Expansion Adaptive Constraints

The proposed shrinkage-expansion adaptive constraints can be flexibly embedded into many pairwise constrained metric learning methods by substituting the original fixed bound based constraints with the adaptive pairwise constraints suggested in Section 3.1. As an example, we embed the shrinkage-expansion adaptive constraints into ITML [6], and call the resulting model SEA-ITML:

$$\begin{aligned} \min_{\mathbf{M}, \xi} D_{ld}(\mathbf{M}, \mathbf{M}_0) + \gamma D_{ld}(\text{diag}(\xi), \text{diag}(\xi_0)) \\ \text{s.t. } D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \leq f_s(D(\mathbf{x}_i, \mathbf{x}_j)) + \xi_{ij} \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{S}, \\ D_{\mathbf{M}}(\mathbf{x}_i, \mathbf{x}_j) \geq f_d(D(\mathbf{x}_i, \mathbf{x}_j)) - \xi_{ij} \quad (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{D}, \\ \xi_{ij} \geq 0, \forall (i, j), \mathbf{M} \succeq 0. \end{aligned} \quad (12)$$

where D_{ld} is the LogDet divergence [16] between matrices.

Apparently, the SEA-ITML model in Eq. (12) is an instantiation of the proposed SEAML framework with the LogDet regularizer. SEA-ITML can be solved using the same algorithm of ITML, and thus SEA-ITML has the same complexity as ITML [6]. It is mentionable that SEA-ITML can achieve much better performance than the original ITML method. Moreover, SEA-ITML performs much better than most of state-of-the-art metric learning methods.

4 A Novel Verification Paradigm with SEAML

Thanks to the proposed shrinkage-expansion rule, the shrinkage-expansion property of the distance before and after metric learning is also an important cue when SEAML is applied to verification. In the training stage, a decision threshold b and a critical

Algorithm 1. Algorithm of Shrinkage-Expansion Adaptive Metric Learning (SEAML)

Input: A set of pairwise training data points $\{(\mathbf{x}_i, \mathbf{x}_j, l_{ij}) | \forall (i, j)\}$ with label l , the trade off parameter C , $\mathbf{Y}_0 = \mathbf{I}$, $t = 0$.

1: **while** Not Converge **do**

2: Fix \mathbf{Y}_t and solve the subproblem Eq. (10) on β^t by SVM solver.

3: Compute $\hat{\mathbf{Y}} = -\sum_{ij} \beta_{ij}^t l_{ij} \mathbf{Z}_{ij}$

4: Fix β^t and solve the subproblem Eq.(11) to update \mathbf{Y}_t : $\mathbf{Y}_t = \mathbf{U}\Sigma_+\mathbf{V}^T$, where $\hat{\mathbf{Y}} = \mathbf{U}\Sigma\mathbf{V}^T$ is the SVD of $\hat{\mathbf{Y}}$ and $\Sigma_+ = \max(\Sigma, \mathbf{0})$.

5: $t = t + 1$.

6: **end while**

7: Obtain the solution $\beta^* = \beta^t$ and $\mathbf{Y}^* = \mathbf{Y}^t$.

8: $\mathbf{M} = \sum_{ij} \beta_{ij}^* l_{ij} \mathbf{Z}_{ij} + \mathbf{Y}^*$.

Output: \mathbf{M}

value δ can be determined from the training data. First, like the traditional verification paradigm, we can find a decision threshold b on the training data using the methods suggested in [10,17,38]. Then, for all training pairs, given decision threshold b and an initialization of δ (e.g., $\delta = 0$), if the absolute value of the difference between the Mahalanobis distance of training pair $(\mathbf{x}_i, \mathbf{x}_j)$ and b is smaller than δ , we adjust its label based on the following paradigm:

$$(\mathbf{x}_i, \mathbf{x}_j) \in \begin{cases} \mathcal{S} & \text{if } D_{ij} - D_{ij}^M \geq 0, \\ \mathcal{D} & \text{if } D_{ij} - D_{ij}^M < 0. \end{cases} \quad (13)$$

where D_{ij} and D_{ij}^M denote the Euclidean distance and the Mahalanobis distance of pair $(\mathbf{x}_i, \mathbf{x}_j)$, respectively. Finally, we can search an optimal solution of δ within a range on the training data. For example, we let $\delta \in [0, 3]$ and set the step size as 0.15.

In the test stage, we take advantage of a two-step strategy to verify if a test pair is similar or dissimilar with the decision threshold b and optimal δ . In the first step, the label of the test pair $(\mathbf{x}_z, \mathbf{x}_k)$ is decided by comparing the Mahalanobis distance with the threshold b . In the second step, if the Mahalanobis distance D_{zk}^M of the pair $(\mathbf{x}_z, \mathbf{x}_k)$ is close to b (i.e., $|D_{zk}^M - b| \leq \delta$), we adjust its label by using Eq. (13). The novel verification paradigm can improve more than 1% the verification performance.

5 Experiments

In this section, we evaluate the proposed method on both classification and face verification problems. The proposed SEAML and SEA-ITML have two tradeoff parameters, i.e., C and γ . Similar to ITML [6] and LMNN [35], these two parameters are tuned via three-fold cross validation in the classification experiments on the UCI datasets. In face verification, we set $C = 1$ and $\gamma = 1$ throughout our experiments. The Matlab code can be downloaded at <http://www4.comp.polyu.edu.hk/~cslzhang/code/SEAML.rar>.

5.1 Classification Results on the UCI Datasets

We first evaluate the classification performances of SEAML and SEA-ITML on 12 datasets from the UCI machine learning repository [2]. These datasets are widely used

to evaluate the classification performance of metric learning methods such as ITML [6], DML-eig [38], etc. Table 1 summarizes the 12 UCI datasets. The k nearest neighbor (kNN) classifier is used for classification, and we set $k = 3$ for all the datasets. We compare SEAML and SEA-ITML with the baseline algorithm without metric learning (i.e., kNN with standard Euclidean distance), and three state-of-the-art metric learning methods: ITML [6], LMNN [35] and DML-eig [38]. The hyper-parameters in ITML and LMNN are tuned via three-fold cross validation, and the classification results of DML-eig are cropped from the original paper [38].

We run the experiments 100 times. In each run, we randomly split each dataset into a training subset and a test subset. The numbers of samples in the training and test subsets are shown in Table 1. The experimental results by averaging over the 100 runs are reported in Table 2. One can see that SEAML and SEA-ITML achieve the best or the second best classification rate on almost every dataset, and they have the best and the second best average rank, outperforming DML-eig, ITML and LMNN.

5.2 Face Verification on LFW

Experimental Setting. In this subsection, we evaluate the proposed SEAML and SEA-ITML methods for face verification on the Labeled Faces in the Wild (LFW) database [15]. LFW is a challenging database designed for studying the unconstrained face verification problem. It contains 13,233 face images from 5,749 persons. Figure 3 shows some samples of similar and dissimilar pairs. There are two commonly used test protocols on LFW: image restricted and image unrestricted protocols. Under the image restricted protocol, the identities of the training images are ignored. The only available information is whether each pair of training images are from the same subject or not. The performance of a face verification algorithm is evaluated by 10-fold cross-validation, and each fold contains a subset of 300 positive and 300 negative image pairs.

To be consistent with the experimental settings used in most of state-of-the-art metric learning methods [10,17,22,38], we adopt the image restricted protocol with the face images aligned by the funneling method [13]. We represent face images by using

Table 1. Summary of the UCI datasets used in the experiment

	Wine	Iris	Segment.	Optdigits	Letter	Sonar	Waveform	Diabetes	Iono.	Breast	Face	Cardi.
# Classes	3	3	7	10	26	2	3	2	2	2	40	10
# Samples for Training	125	105	1617	1934	16000	188	3500	538	176	398	280	1914
# Samples for Test	53	45	693	946	4000	20	1500	230	175	171	120	212
Dimension	13	4	19	64	16	60	21	8	34	30	64	21

Table 2. Classification rates (%) on the UCI databases. The best and second best results on each dataset are highlighted in red and blue, respectively.

Method	Wine	Iris	Segment.	Optdigits	Letter	Sonar	Waveform	Diabetes	Iono.	Breast	Face	Cardi.	Ave. Rank
Baseline	94.24	94.89	94.28	98.33	95.54	79.58	81.13	68.91	81.73	93.53	93.33	75.34	5.92
ITML	96.18	95.56	94.98	98.54	97.22	85.14	84.06	70.04	85.52	93.18	97.58	80.1	4.67
LMNN	96.92	95.78	96.31	98.63	96.55	88.43	81.39	70.3	87.41	94.65	97.92	80.79	3.17
DML-eig	98.65	96.89	97.03	98.55	96.14	86.5	84.67	72.29	83.74	96.47	98.33	79.71	2.83
SEA-ITML	97.79	96.94	96.87	98.74	97.81	87.97	84.54	71.56	86.67	96.37	97.62	80.37	2.5
SEAML	98.54	97.13	97.21	98.67	97.77	88.67	85.02	71.94	86.81	95.75	97.89	80.31	1.92

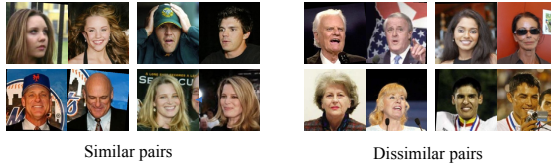


Fig. 3. Some examples of face image pairs in the LFW face database [15]. Left: four similar pairs; Right: four dissimilar pairs.

their SIFT features [25] provided by [10], where 128 dimensional SIFT descriptors are extracted on three scales and at 9 fiducial points (corners of the mouth, eyes, and nose) detected by a facial feature detector. Finally, each face is represented by a 3,456 dimensional SIFT feature vector. This dimensionality is rather too high for stable metric learning. Therefore, to be consistent with [10,38], principal component analysis (PCA) is applied to project each SIFT feature vector into a d -dimensional subspace. Then a $d \times d$ PSD matrix is learned as the metric matrix for face verification.

Dimension of Principal Components. The dimension d of principal components has an important impact on the verification performance of metric learning algorithms. Figure 4 shows the verification accuracy of SEAML, ITML [6], LDML [10] and DML-eig [38] with various dimensions of principal components. We can see that SEAML performs the best under all dimensions, while DML-eig performs the second best. With the increase of dimension, ITML, LDML, and DML-eig achieve their best results at dimension 55, 35, and 100, respectively. Their performance will then drop rapidly. Unlike the other methods, the verification accuracy of the proposed SEAML increases stably with the increase of PCA dimension until 170, and its verification accuracy only drops slightly (about 0.6%) when d is 200. In summary, SEAML has not only better verification accuracy than the competing methods, but also more robustness to the feature dimension. This validates that the proposed shrinkage-expansion adaptive constraints are effective in improving the accuracy and stability of metric learning.

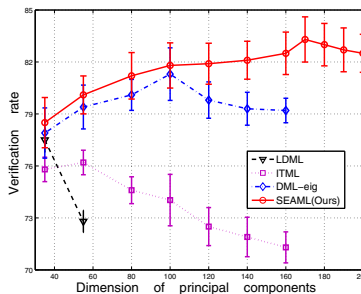


Fig. 4. Comparison of LDML, ITML, DML-eig and SEAML in face verification under different dimensions of SIFT features

Comparison with State-of-the-Arts. We compare SEAML with 13 state-of-the-art methods [1,6,10,14,17,19,20,22,23,28,32,36,38] under the same protocol of image restricted setting. Note that LMNN is not compared because it requires triplet constraints which are not applicable to the LFW database under restricted setting.

In Table 3, we first compare SEAML with those metric learning based methods using the SIFT features. Clearly, SEAML and SEA-ITML outperform all the other metric learning methods, validating the effectiveness of the proposed shrinkage-expansion adaptive pairwise constraints. Besides SIFT features, we also evaluate SEAML using the attribute features provided by [19] (please refer to Section 5.3 for more information of attribute features). With the attribute features, SEAML and SEA-ITML achieve a similar verification rate of 85.8%, outperforming state-of-the-art metric learning methods, such as LDML [10](83.4%), ITML [6] (84.0%) and KissMe [17] (84.6%). SEAML and SEA-ITML also perform slightly better than the Attribute and Simile classifiers (85.3%) [19], which achieves state-of-the-art performance using the SVM based method.

In Table 4, we evaluate SEAML by combining the SIFT and Attribute features, and compare it with those state-of-the-art non-metric learning methods or metric learning methods with multiple features. We can see that the verification accuracy (87.5%) of SEAML is the second highest, the same as Fisher vector face (87.5%) [32]. Kindly note that [32] uses much more complicated strategies, including Fisher vector based face representation and discriminative dimensionality reduction, joint distance-similarity metric learning and horizontal flipping. The method in [23] reports the best accuracy (89.6%). This method employs a learned second-order decision functions solved by kernel SVM (may suffer from high memory burden in large scale problem) for verification and it combines SIFT with complex hierarchical Gaussianization (HG) vector features [39].

Table 3. Verification accuracies (%) of competing metric learning methods on the LFW-funneled database under the image restricted protocol. (Symbol “-” means that the result is not available.)

Metric learning with single feature		
Method	SIFT	Attributes
LDML [10]	77.5 ± 0.5	83.4
ITML [10]	76.2 ± 0.5	84.0
LDA-based [36]	79.4 ± 0.2	-
DML-eig [38]	81.3 ± 2.3	-
SODFML [22]	81.0	-
KissMe [17]	80.5	84.6
SEA-ITML	82.4 ± 1.7	85.8
SEAML	83.2 ± 1.2	85.8

Table 4. Comparison with state-of-the-art results on the LFW-funneled database under the image restricted protocol

Method	Features	Accuracy (%)
MERL+Nowark [28]	SIFT and geometry	76.2 ± 0.6
V1-like/MKL [14]	V1-like features	79.4 ± 0.5
MRF-MLBP [1]	Multi-scale LBP	79.1 ± 0.2
APEM(fusion) [20]	Dense SIFT + LBP	84.1 ± 1.2
Attr.& sim. classifiers [19]	Attributes	85.3
DML-eig combined [38]	SIFT + LBP + TPLBP + FPLBP	85.7 ± 0.6
Fisher vector face [32]	Dense SIFT	87.5 ± 1.5
Li et al. [23]	SIFT + HG vector features	89.6
SEA-ITML	SIFT + Attributes	87.2 ± 1.7
SEAML	SIFT + Attributes	87.5 ± 1.3

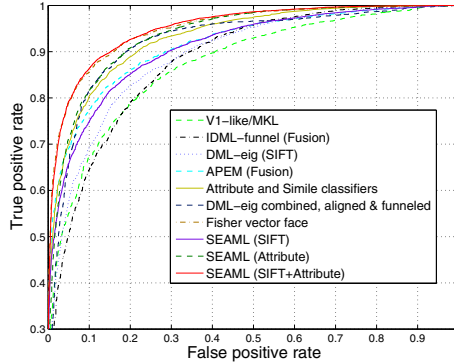


Fig. 5. ROC curves of various methods on the restricted LFW-funneled database

In Figure 5, we plot the ROC curves of various methods on the restricted LFW-funneled dataset. We can see that SEAML and Fisher vector face show comparable performance and they outperform the other competing methods. The experimental results clearly demonstrated that the proposed SEAML method can learn effective metrics from training data with complex structures and distributions.

5.3 Face Verification on PubFig

Experimental Setting. The Public Figures (PubFig) [19] face database shares similar philosophy with LFW. It is also a large, real-world face database collected from the internet, consisting of 58,797 images from 200 people. PubFig is a very challenging database with large face variations in pose, expression, lighting, scene, camera, and so on. Like LFW, PubFig provides an evaluation platform for face verification: 20,000 pairs of images from 140 people are divided into 10 cross-validation folds, and each fold consists of 1,000 intra and 1,000 extra pairs from disjoint sets of 14 people. Some illustrative examples of similar and dissimilar pairs are shown in Figure 6.

Unlike most of the existing face databases which use handcrafted features to represent face images, PubFig provides high-level semantic features which contain 73 kinds of attributes, such as hair, glass, age, race, smiling and so on. All attributes are computed by the effective attribute classifiers [19]. These attribute features allow the elaboration of semantic description which are more robust against large image variations, and can lead to good verification performance. In this section, to be consistent with [6,10,17,19,35], we use these high-level semantic features to evaluate SEAML and SEA-ITML.

Comparison with State-of-the-Arts. Apart from the baseline with Euclidean distance, we compare SEAML and SEA-ITML with ITML [6], LMNN [35], LDML [10], KissMe [17] and SVM [19]. The results of these competing methods are obtained from [17], and they are all metric learning based methods except for SVM. Kindly note that the publically reported results on the PubFig dataset are not as many as those on the LFW dataset, and the compared results here are the best ones in literature.

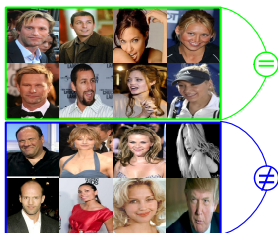


Fig. 6. Examples of similar and dissimilar pairs in the PubFig face database [19]

The verification results (Equal Error Rate, EER) of the competing methods are listed in Table 5. One can see that SEAML and SEA-ITML are the best two methods. LDML, KissMe and SVM also perform well. ITML does not perform well on this dataset and its EER (69.3%) is even worse than the baseline with Euclidean distance. However, by adopting the proposed shrinkage-expansion adaptive constraints into ITML, the resulting SEA-ITML method can achieve an EER of 77.8%. This further validates the effectiveness of the proposed shrinkage-expansion adaptive constraints in improving the verification performance of pairwise constrained metric learning methods.

Table 5. Verification performance on the PubFig face database. All methods use the attribute features provided by [19]. The reported results of all compared methods are copied from [17].

Method	EER
L_2 (Euclidean distance)	72.5%
ITML	69.3%
LMNN	73.5%
LDML	77.6%
KissMe	77.6%
SVM	77.6%
SEA-ITML	77.8%
SEAML	78.6%

6 Conclusion

In this paper, we presented a novel metric learning framework, namely shrinkage expansion adaptive metric learning (SEAML). With the proposed shrinkage-expansion adaptive pairwise constraints, SEAML can learn more effectively the distance metrics from data with complex distributions, and endow a new paradigm to determine whether a pair is matched or not by considering both the decision threshold and the change of matching distances before and after metric learning. Moreover, the proposed shrinkage-expansion rule can be embedded into many existing pairwise constrained metric learning methods (e.g., ITML [6]), and improve much their performance. Experimental results on the UCI datasets, the LFW and PubFig face databases validated that SEAML outperforms many state-of-the-art metric learning algorithms. In the future work, we will investigate

how to adopt a low rank regularizer [37,24] in the SEAML framework to further improve its robustness against the feature dimension and computational efficiency. We are also interested in studying a flexible scheme which can learn optimized parameters of proposed shrinkage-expansion adaptive constraints.

Acknowledgements. This work is supported by the Hong Kong RGC GRF grant (PolyU 5313/13E), NSFC grants (61271093, 60973080, 61170149), the program of MoE (NCET-12-0150), and the Fundamental Research Funds for the Central Universities (DUT13RC(3)02).

References

1. Arashloo, S.R., Kittler, J.: Efficient processing of MRFs for unconstrained-pose face recognition. In: BTAS (2013)
2. Asuncion, A., Newman, D.: UCI machine learning repository (2007), <http://www.ics.uci.edu/~mllearn/MLRepository.html>
3. Bellet, A., Habrard, A., Sebban, M.: A survey on metric learning for feature vectors and structured data. arXiv 1306.6709 (2013)
4. Chang, C.C., Lin, C.J.: Libsvm: A library for support vector machines. ACM Trans. Intell. Syst. Technol. 2(3), 1–27 (2011)
5. Csiszár, I., Tushnádý, G.: Information geometry and alternating minimization procedures. Statistics and Decisions (1984)
6. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: ICML (2007)
7. Feng, Z., Jin, R., Jain, A.: Large-scale image annotation by efficient and robust kernel metric learning. In: ICCV (2013)
8. Globerson, A., Roweis, S.: Metric learning by collapsing classes. In: NIPS (2005)
9. Goldberger, J., Roweis, S., Hinton, G., Salakhutdinov, R.: Neighbourhood components analysis. In: NIPS (2004)
10. Guillaumin, M., Verbeek, J., Schmid, C.: Is that you? metric learning approaches for face identification. In: ICCV (2009)
11. Huang, C., Zhu, S., Yu, K.: Large scale strongly supervised ensemble metric learning, with applications to face verification and retrieval. arXiv 1212.6094 (2012)
12. Huang, D., Storer, M., De la Torre, F., Bischof, H.: Supervised local subspace learning for continuous head pose estimation. In: CVPR (2011)
13. Huang, G.B., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. In: ICCV (2007)
14. Huang, G.B., Jones, M.J., Learned-Miller, E.: LFW results using a combined Nowak plus MERL recognizer. In: Real-Life Images Workshop at the ECCV (2008)
15. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Tech. rep., University of Massachusetts, Amherst (2007)
16. James, W., Stein, J.: Estimation with quadratic loss. In: Berkeley Symposium on Mathematical Statistics and Probability (1961)
17. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: CVPR (2012)
18. Kulis, B., Saenko, K., Darrell, T.: What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In: CVPR (2011)

19. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: ICCV (2009)
20. Li, H., Hua, G., Lin, Z., Brandt, J., Yang, J.: Probabilistic elastic matching for pose variant face verification. In: CVPR (2013)
21. Li, X., Shen, C., Shi, Q., Dick, A., van den Hengel, A.: Non-sparse linear representations for visual tracking with online reservoir metric learning. In: CVPR (2012)
22. Li, Z., Cao, L., Chang, S., Smith, J.R., Huang, T.S.: Beyond mahalnobis distance: Learning second-order discriminant function for people verification. In: CVPR Workshops (2012)
23. Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R.: Learning locally-adaptive decision functions for person verification. In: CVPR (2013)
24. Lim, D., Mcfee, B., Lanckriet, G.R.: Robust structural metric learning. In: ICML (2013)
25. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* 60(2), 91–110 (2004)
26. Mensink, T., Verbeek, J., Perronnin, F., Csurka, G.: Metric learning for large scale image classification: Generalizing to new classes at near-zero cost. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 488–501. Springer, Heidelberg (2012)
27. Mignon, A., Jurie, F.: Pcca: A new approach for distance learning from sparse pairwise constraints. In: CVPR (2012)
28. Nowak, E., Jurie, F.: Learning visual similarity measures for comparing never seen objects. In: CVPR (2007)
29. Parameswaran, S., Weinberger, K.: Large margin multi-task metric learning. In: NIPS (2010)
30. Shen, C., Kim, J., Wang, L.: A scalable dual approach to semidefinite metric learning. In: CVPR (2011)
31. Shen, C., Kim, J., Wang, L., van den Hengel, A.: Positive semidefinite metric learning with boosting. In: NIPS (2009)
32. Simonyan, K., Parkhi, O.M., Vedaldi, A., Zisserman, A.: Fisher Vector Faces in the Wild. In: BMVC (2013)
33. Verma, Y., Jawahar, C.V.: Image annotation using metric learning in semantic neighbourhoods. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 836–849. Springer, Heidelberg (2012)
34. Weinberger, K.Q., Saul, L.K.: Fast solvers and efficient implementations for distance metric learning. In: ICML (2008)
35. Weinberger, K., Blitzer, J., Saul, L.: Distance metric learning for large margin nearest neighbor classification. In: NIPS (2006)
36. Wolf, L., Hassner, T., Taigman, Y.: Descriptor based methods in the wild. In: Real-Life Images Workshop at the ECCV (2008)
37. Ying, Y., Huang, K., Campbell, C.: Sparse metric learning via smooth optimization. In: NIPS (2009)
38. Ying, Y., Li, P.: Distance metric learning with eigenvalue optimization. *J. Mach. Learn. Res.* 13(1), 1–26 (2012)
39. Zhou, X., Cui, N., Li, Z., Liang, F., Huang, T.S.: Hierarchical gaussianization for image classification. In: ICCV (2009)