# Maximum-Margin Coupled Mappings for Cross-Domain Face Recognition

Stephen Siena, *Student Member, IEEE,* Vishnu Naresh Boddeti, *Member, IEEE,* and B.V.K. Vijaya Kumar, *Fellow, IEEE*

***Abstract*—Over the past decade, fueled by cheaper storage and availability of ever increasing computational resources, there has been an explosive increase in the collection of data about the same concept (e.g. face images) from multiple sources and in multiple formats. This leads to pattern matching scenarios across that necessitate the development of learning algorithms that have the ability to learn concepts from diverse sources of data, much like human learning. In this paper, we consider this problem of concept learning from and across multiple sources of data in the context of face recognition under challenging scenarios i.e., we consider situations in the wild where the probe and gallery images are captured under different conditions, thereby treating the probe and gallery images as arising from different domains. We present Maximum-Margin Coupled Mappings (MMCM), a method which learns linear projections to map the data from the different domains into a common latent domain, maximizing the margin of separation between between the intraclass data and the interclass data from different domains. We demonstrate the effectiveness of this technique on multiple face biometric datasets with a variety of cross-domain relationships.**

***Index Terms*—Biometrics, Face Recognition, Coupled-Mappings, Super-Resolution**

## I. Introduction

FACE recognition has seen tremendous progress over the past decade, leading to excellent performance under controlled conditions. However, face recognition still remains a very challenging problem under uncontrolled conditions due to factors such as pose, illumination, expression, image resolutions, different camera sensors, etc. Efforts have been made to address these challenging factors in face recognition [1], [2], [3], [4], [5]. In this paper we consider the problem of matching facial images across two different data domains i.e., across image resolutions and camera sensors. While traditional machine learning algorithms for face recognition have largely focused on concept learning from a single data source (e.g., Principal Component Analysis [6], Linear Discriminant Analysis [7], etc.), matching facial images across different data domains necessitates the development of learning algorithms using data from multiple domains.

More specifically, in the context of biometric image matching, the focus of machine learning research has been to perform matching between a probe sample and a gallery set of data for tasks such as face verification and recognition. Oftentimes the probe lies in the same domain as the gallery set, but this is not always the case. For example, in some face recognition scenarios, the probe images captured in uncontrolled settings (e.g. surveillance cameras) at a lower resolution than the set of gallery images. Also, probe images may be captured with one type of sensor while the gallery images may have been acquired using a different type of sensor. When the probe image is significantly different than the gallery, classifiers trained on the gallery may not be able to treat the probe image in a way that allows for an effective comparison. In cases where the gallery and probe images have fundamentally different feature representations, methods effective for matching tasks in a single domain may be completely impractical.

The techniques for handling data from different domains can depend on the objectives . *Domain adaptation* and *transfer learning* refer to approaches in settings where one domain often has significantly more training data, and the goal is to learn a mapping from one domain to the other. In contrast, *coupled mappings* focus on learning mappings to a latent subspace common to both domains, typically leveraging equal amounts of data from each domain. Additionally, this subspace can be low-dimensional, in which case coupled mapping formulations also perform dimensionality reduction with the assumption that the underlying data lies in a latent low-dimensional subspace that both domains can be mapped to. Many different coupled-mapping algorithms for cross-domain face recognition have been proposed, especially in the context of matching low-resolution images against high-resolution images, where the images are matched using a nearest neighbor classifier in the common, possibly low-dimensional, subspace.

In this paper we present a new coupled mapping formulation called Maximum-Margin Coupled Mappings (MMCM), which combines the common subspace learning principle of coupled mapping techniques and the margin maximizing properties of single domain large margin nearest neighbor [8] methods. The margin maximizing formulation of the proposed coupled-mapping algorithm helps improve the robustness of the algorithm to noise and outliers in comparison to other coupled-mapping algorithms. While the existing algorithms learn the linear projections to maximize the average distance between the authentic and impostor match pairs, the proposed method learns linear projections to maximize the distance between the closest (worst case example) authentic and impostor match pairs, thereby increasing the robustness to noise.

A preliminary version of the algorithm presented in this paper was introduced in [4]. In this paper we present a more in-depth analysis of MMCM along with experimental evaluation under multiple cross-domain face matching scenarios. More

The authors are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213. E-mail: ssiena@andrew.cmu.edu, naresh@cmu.edu, kumar@ece.cmu.edu

specifically, the main contributions of this paper are:

1) The basic MMCM formulation and an in-depth analysis of the technique discussing the intuition behind the development of MMCM relative to other related techniques.
2) We also discuss how to optimize MMCM efficiently with a new fast gradient descent algorithm, making MMCM's application to larger datasets more practical.
3) Lastly, we present an extensive evaluation of MMCM under different cross-domain matching scenarios. We show results on problems that model both identification and verification biometric problems, and show results on a range of face recognition datasets covering a range of cross-domain problems. Additionally, we incorporate a simple face detector into our evaluations to minimize the need for manually labeled facial landmarks. These experimental procedures provide an illustration of the strengths and the weaknesses of our proposed algorithm.

The paper is structured as follows: Section II defines the problem coupled mappings aim to solve. Section III discusses related work, particularly those related to coupled mappings. Section IV discusses how MMCM is designed to solve cross-domain matching problems. Section V introduces the datasets, and the results are discussed in Section VI. Section VII discusses additional properties of MMCM. Section VIII concludes the paper.

## II. Problem Statement

In many settings, data samples are collected from two different domains, $\mathcal{D}_A \in \mathbb{R}^\alpha$ and $\mathcal{D}_B \in \mathbb{R}^\beta$, which share a common set of labels $\mathcal{L} = \{l_1, l_2, ..., l_c\}$. If we are given $\mathbf{A} = \{a_1, a_2, ..., a_m\} \in \mathcal{D}_A$ and $\mathbf{B} = \{b_1, b_2, ..., b_n\} \in \mathcal{D}_B$, the goal is to match samples from one domain to samples from the other. Defining a distance metric is relatively straightforward for data samples from the same domain and the same feature representation, but this may not be the case when comparing data across different domains. To facilitate cross-domain comparisons, we map data from $\mathcal{D}_A$ and $\mathcal{D}_B$ into a common space, which is denoted as $\mathcal{D}_Z \in \mathbb{R}^\gamma$. The task is to learn mappings $f_A : \mathcal{D}_A \rightarrow \mathcal{D}_Z$ and $f_B : \mathcal{D}_B \rightarrow \mathcal{D}_Z$. MMCM is a method of learning linear projections, therefore $f_A : \mathbf{P}_A^T a_i \rightarrow \hat{a}_i$ and $f_B : \mathbf{P}_B^T b_i \rightarrow \hat{b}_i$. $\mathbf{P}_A \in \mathbb{R}^{\alpha \times \gamma}$ projects the $\alpha$-dimensional data sample $a_i$ into a $\gamma$-dimensional subspace. Similarly, $\mathbf{P}_B \in \mathbb{R}^{\beta \times \gamma}$ projects the $\beta$-dimensional data sample $b_i$ into a $\gamma$-dimensional subspace. Once $\mathbf{A}$ and $\mathbf{B}$ are projected to $\hat{\mathbf{A}} = \{\hat{a}_1, \hat{a}_2, ..., \hat{a}_m\} \in \mathcal{D}_Z$ and $\hat{\mathbf{B}} = \{\hat{b}_1, \hat{b}_2, ..., \hat{b}_m\} \in \mathcal{D}_Z$, distances between data samples from the two different domains can be more easily compared. Typically simple metrics such as Euclidean distance (or some other distance metric) can be used to find the similarity between samples from the different domains. When given class labels $Y_A = \{y_{a1}, y_{a2}, ..., y_{am}\} \in \mathcal{L}$ and $Y_B = \{y_{b1}, y_{b2}, ..., y_{bn}\} \in \mathcal{L}$, ideally we will learn projections such that $d_{y_{ai}=y_{bj}}(\hat{a}_i, \hat{b}_j) < d_{y_{ai} \neq y_{bj}}(\hat{a}_i, \hat{b}_k)$, where $d(\cdot, \cdot)$ is some distance metric. In other words, given data samples from $\mathcal{D}_A$ and data samples from $\mathcal{D}_B$, we want the cross-domain match pairs to be closer together than the cross-domain non-match pairs.

## III. Related Work

The main premise of this line of research is that, *there exists a functional relation between the different sources of data since they are generated from the same underlying latent phenomenon*. This latent source is obvious in some scenarios (like when the same data is captured by different sensors, different viewpoints etc.) and not so in other scenarios (like the relation between the text in a news article and the image of a subject's face in that article). Assuming that the premise holds, the goal of the learning algorithms is to estimate or learn the function, relating the different data sources, either exactly or approximately. Many algorithms exist and can be designed for this purpose. We note general paradigms and specific algorithms for doing this. At a high level each of these algorithms are designed to learn features from each data source such that similar concepts are mapped close to each other.

An increasing range of applications leverage domain adaptation and transfer learning methods. It is used in natural language processing [9], as well as computer vision tasks [10], [11], as well as multimodal learning tasks [12], [13]. Pan & Yang offer a more complete survey of these fields [14].

Within biometrics, there has also been work addressing cross-domain challenges. Hennings-Yeomans et al. proposed a method for comparing low-resolution face images to a high-resolution gallery [3]. Klare and Jain introduced a framework for heterogeneous face recognition using prototypes in multiple imaging modalities to compare different image representations [5]. Zuo et al. perform cross-spectral iris recognition by learning a predictive mapping function from the visible light spectrum to the near infrared channel [15]. Pillai et al. learn a kernel transformation and impose sensor adaptation constraints for cross-sensor iris matching[16].

### A. Coupled Mappings

Coupled mappings aim to learn projections to a single subspace, with the assumption that data from multiple sources actually lie on a subspace whose dimensionality is lower than the data sources. These approaches seek to take advantage of any correlation between the data sources to learn functions that map different data sources close to each other in the mapped space, and work in conjunction with non-parametric classifiers like nearest neighbor matching.

Canonical Correlation Analysis (CCA) learns projections which maximize the correlation between two sets of data [17]. Li et al. introduced two methods, one referred to as Coupled Mappings, and an augmented technique called Coupled Locality Preserving Mappings (CLPM), which preserves local neighborhoods of data samples when learning the mappings [18]. Yang et al. added local consistency terms to preserve distances within each domain, and show improvements over CLPM [19]. Kan et al. introduced Multi-view Discriminant Analysis (MvDA), which was proposed to potentially learn more than two sets of projections from any number of domains to a single common subspace [20].

*1) Simultaneous Discriminant Analysis:* Simultaneous Discriminant Analysis (SDA) was introduced for the purpose of matching low resolution (LR) face images to high resolution (HR) face images [21]. It can be seen as a cross-domain extension of Linear Discriminant Analysis (LDA) [7]. While LDA learns projections to minimize intraclass variation and maximize interclass separation in the learned subspace for one domain, SDA optimizes the same criteria for data across two domains simultaneously. Intraclass variation is defined as the sample variance within each class in the projected space, and interclass variation is defined as the sample between-class covariance in the projected space.

To learn projections that improve both the intraclass variation and the interclass variation, the full objective can be written as a ratio of the two; so the objective for SDA is

$$\arg\max_{P} \frac{P^T \mathbf{S}_b P}{P^T \mathbf{S}_w P}. \tag{1}$$

where $P^T \mathbf{S}_w P$ and $P^T \mathbf{S}_b P$ capture the intraclass and interclass variation of the data from both domains, respectively, and $P = [P_A^T P_B^T]^T$. This maximization results in a generalized eigenvalue problem (GEV)

$$\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{P} = \mathbf{P} \Lambda, \tag{2}$$

where the eigenvectors $\mathbf{P}$ corresponding to the $\gamma$-largest eigenvalues comprise the projections $\mathbf{P}_A$ and $\mathbf{P}_B$.

For $\mathbf{S}_w^{-1}$ to be symmetric (required for $\mathbf{S}_w^{-1}\mathbf{S}_b$ to produce real eigenvectors), the number of training samples from each domain must be equal for each class. This constraint is similar to that of CCA. By using the class means and variances, SDA is implicitly modeling the distribution of each class as a Gaussian. How well this assumption holds will depend on the particular application.

*2) Coupled Marginal Fisher Analysis:* While SDA could be seen as a cross-domain extension of LDA, Coupled Marginal Fisher Analysis (CMFA) [22] can be seen as the coupled mapping extension of Marginal Fisher Analysis (MFA) [23]. MFA learns projections that optimize distances in the learned subspace based on the local neighborhoods of the original data. The objective function for CMFA is

$$\min_{\mathbf{P}_A, \mathbf{P}_B} \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} \|\mathbf{P}_A a_i - \mathbf{P}_B b_j\|_2^2 w_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} \|\mathbf{P}_A a_i - \mathbf{P}_B b_j\|_2^2 w_{ij}^P}, \tag{3}$$

where $w_{ij} \in \mathbf{W}$ and $w_{ij}^P \in \mathbf{W}^P$, where $\mathbf{W} \in \mathbb{R}^{m \times n}$ and $\mathbf{W}^P \in \mathbb{R}^{m \times n}$ represent connectivity graphs between data samples. $w_{ij} \in \mathbf{W}$ represents how strongly the distance between a pair of cross-domain data samples should be minimized (corresponding to data samples in the same class), and $w_{ij}^P \in \mathbf{W}^P$ represents how strongly the distance between a pair of cross-domain data samples should be maximized (corresponding to data samples from different classes). In CMFA, $w_{ij}$ is non-zero if $a_i$ and $a_j$ have the same class label and are in the same local region, according to some number of nearest neighbors. Similarly, $w_{ij}^P$ is non-zero if $a_i$ and $a_j$ have different class labels and are in the same local region according to some number of nearest neighbors.

Note that while the objective function optimizes cross-domain distances, $\mathbf{W}$ and $\mathbf{W}^P$ often must be defined by comparisons of data samples in a single domain. This is because when the domains do not have feature representations of the same dimension, the cross-domain distance is not well-defined before being projected.

Similar to SDA, CMFA can be reformulated into a GEV. While SDA is best suited for a Gaussian distribution of each class, CMFA relies on optimizing the local neighborhoods of data samples. CMFA was shown to outperform SDA in a LR face recognition task[22].

*3) Multi-Dimensional Scaling:* Biswas et al. take a multi-dimensional scaling (MDS) approach to LR face recognition [24]. The objective is

$$\begin{aligned}\min_{\mathbf{P}_A, \mathbf{P}_B} \lambda \sum_{i=1}^{m}\sum_{j=1}^{n} (d(\mathbf{P}_A^T a_i, \mathbf{P}_B^T b_j) - d(\mathbf{P}_B^T b_i, \mathbf{P}_B^T b_j)) \\ + (1-\lambda) \sum_{i=1}^{m}\sum_{j=1}^{n} \delta(y_i, y_j) d(\mathbf{P}_A^T a_i, \mathbf{P}_B^T b_j),\end{aligned} \tag{4}$$

where $\lambda$ is a tradeoff parameter, and $\delta(y_i, y_j) = 1$ if $y_i = y_j$, and 0 otherwise (with the assumption that $y_{a_i} = y_{b_i}$). The assumption of the first term, the *distance preserving* term, is that given two HR face images, the distances between matching images will be small, and the distance between non-matching images will be large. With that in mind, the MDS approach learns projections such that the distance between a LR and HR image approximates the distance of two corresponding HR images. Additionally, the second term of the objective, the *class separation* term, penalizes large distances between LR and HR images from the same class. Biswas et al. show how the objective can be reformulated to a form that can be solved by iterative majorization [25]. Iterative majorization solves a closely related function iteratively until some stopping criterion is met. Biswas et al. introduced the MDS method for LR face recognition, and because face recognition is more effective at higher resolutions, the objective is sensible. However, if the domains and application are different, learning projections to mimic the distances of data in one domain may not be effective.

With the exception of CCA and MvDA, all the coupled mapping methods discussed were introduced for the purpose of low-resolution face recognition. Much of the focus of coupled mappings is to leverage the information available in high-resolution face images to help recognize low-resolution face images. While this has traditionally been the purpose of the previous coupled mapping methods, we intend to show the utility of these approaches for both low resolution face recognition and heterogeneous face recognition, where images from different imaging modalities are matched.

## IV. Methodology

Coupled mapping methods focus on learning mappings to a latent subspace common to both domains. In this section we describe our proposed coupled mapping technique, MMCM. Figure 1 gives a pictorial overview showing how MMCM maps data from $\mathcal{D}_A$ and $\mathcal{D}_B$ to a common subspace $\mathcal{D}_Z$ such that there is an explicit margin of separation of between the match and non-match pairs. We also note that the mappings

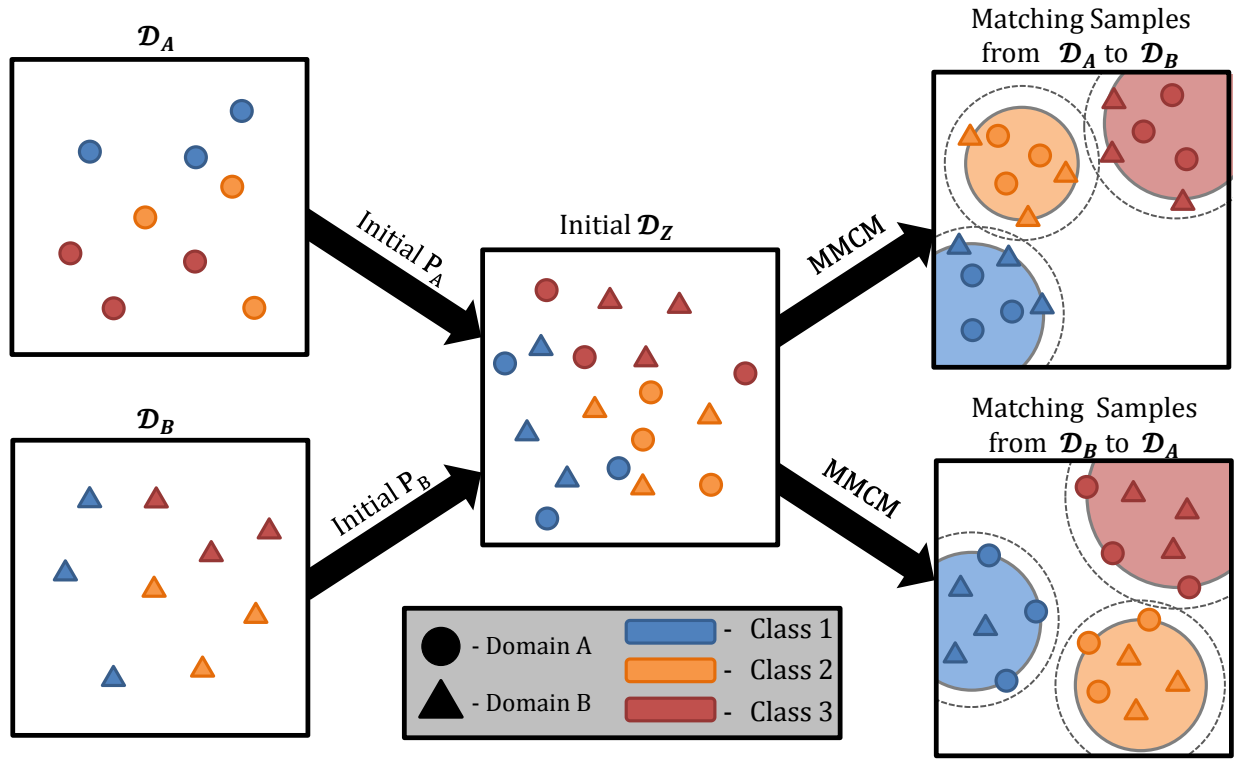



Fig. 1: Overview of coupled mappings and MMCM. Coupled mappings project data from $\mathcal{D}_A$ and $\mathcal{D}_B$ into a common subspace, $\mathcal{D}_Z$, where matching can be done between data samples from the different domains. MMCM learns mappings optimized for matching data samples from one domain to the second, but not vice versa (i.e. comparing samples from $\mathcal{D}_A$ to a gallery from $\mathcal{D}_B$, but not samples from $\mathcal{D}_B$ to a gallery from $\mathcal{D}_A$, or vice versa). MMCM learns coupled mappings such that there is a margin between cross-domain matches and the nearest cross-domain non-matches. The cross-domain match distances define a perimeter around each class, and no cross-domain non-matches enter a margin extending from this perimeter.

learned by MMCM are different depending on the whether the data samples are being matched from one domain to the other or vice-versa. Just like most other coupled mapping methods can be seen as extensions of single domain methods, MMCM is inspired by Large Margin Nearest Neighbor (LMNN) classification [8]. LMNN learns a Mahalanobis distance such that same class pairs have a small distance while ensuring a margin between pairs of data from different classes.

### A. MMCM Formulation

MMCM is inspired by large margin based distance learning methods [26], [8] which have demonstrated success in single domain problems. We learn the projections $\mathbf{P}_A$ and $\mathbf{P}_B$ by solving the following minimization problem:

$$\min_{\mathbf{P}_A,\mathbf{P}_B} \lambda f_{pull}(\mathbf{P}_A,\mathbf{P}_B) + (1-\lambda) f_{push}(\mathbf{P}_A,\mathbf{P}_B) \quad (5)$$

where expanding $f_{pull}$ and $f_{push}$ are defined as

$$f_{pull} = \frac{\sum_M d(a_i,b_j)}{|M|} \quad (6)$$

$$f_{push} = \frac{\sum_V [1 + d(a_i,b_j) - d(a_i,b_k)]_+}{|V|} \quad (7)$$

where $M = \{(i,j)|y_{ai} = y_{bj}\}$ and $V = \{(i,j,k)|y_{ai} = y_{bj}, y_{ai} \neq y_{bk}\}$ and $[x]_+$ signifies the hinge loss, defined as $max(0,x)$. In practice we match test samples to our gallery using Euclidean distance nearest neighbor, so we use the squared Euclidean distance to compare data samples, thus

$$d(a_i,b_j) = \|\mathbf{P}_A^T a_i - \mathbf{P}_B^T b_j\|_2^2. \quad (8)$$

Equation 5 contains two terms, with $\lambda = (0,1)$ acting as the tradeoff parameter between the two ($|M|$ and $|V|$ also normalize the scale of the two terms). Figure 2 illustrates how the two components of MMCM work in conjunction to learn projections. Minimizing the first term, $f_{pull}$, brings cross-domain match pairs close together. The second term, $f_{push}$, consists of a margin, set to 1, a cross-domain match pair, and a cross-domain non-match pair. In reducing the hinge loss of these three components to zero, the non-match $b_k$ is farther from $a_i$ compared to $b_j$ by at least a margin of 1. It is important to note that even in the case where data is linearly separable, even though the margin is a fixed value in the minimization, $|\mathbf{P}_A|$ and $|\mathbf{P}_B|$ cannot simply approach $\infty$ (thus ensuring there is always a margin of at least 1). As $|\mathbf{P}_A|$ and $|\mathbf{P}_B|$ increase, so too will $f_{pull}$. Similarly, $|\mathbf{P}_A| \to 0$ and $|\mathbf{P}_B| \to 0$, the fixed margin of 1 will be violated in an increasing number of triplets in $V$. Thus, $f_{pull}$ and $f_{push}$ work antagonistically to ensure that solving the minimization does not result in a trivial solution, but rather mappings that result in tightly clustered

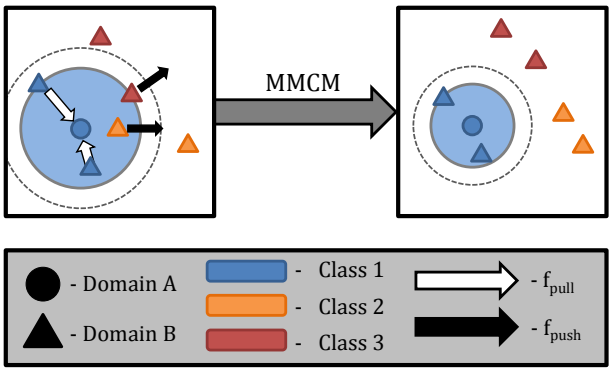



Fig. 2: Illustration showing the objective of the MMCM optimization, and the effect of $f_{pull}$ and $f_{push}$ (for simplicity, only a single data sample from $\mathcal{D}_A$ is shown). The large blue circle represents the perimeter defined by the largest match pair distance, and the dashed line shows the margin extending from that perimeter. Initially, cross-domain matches are far from the data sample from $\mathcal{D}_A$, and cross-domain non-matches are within boundary of the margin. The $f_{pull}$ term brings together the match pairs, and $f_{push}$ moves the intruding non-matches outside of the boundary of the margin, while non-matches outside of the margin are not explicitly acted upon. The result is that the data samples from Class 1 are closer together, and non-matches are now outside of the margin around Class 1.

same class data that has large separation from other classes. Learning a margin of separation between classes has benefits over other coupled mappings. The hinge loss associated with margin learning ensures that the worst cases seen in training have the largest impact on the learned projections. Other coupled mappings, such as SDA and MDS, learn projections that optimize the average case. If interclass samples are already well separated, there is no need to push them farther apart. Our margin based technique instead learns projections to handle the most challenging cases seen in training without further optimizing adequately separated data.

It is important to note that $f_{push}$ does not treat data samples from $\mathcal{D}_A$ the same as data samples from $\mathcal{D}_B$. Specifically, the margins are only explicitly learned for data samples from $\mathcal{D}_A$ with respect to match pairs and non-match pairs from $\mathcal{D}_B$. By only optimizing the margins for data from one domain, MMCM focuses only on matching from one domain to the other. For example, in LR face recognition, MMCM will optimize matching LR images to a gallery of HR images, while not explicitly learning to match HR images to a gallery of LR images. Other coupled mapping techniques such as SDA, and CMFA, will balance the two cross-domain matching cases, even when an application does not need it. In a case where it may be desirable to match from $\mathcal{D}_A$ to $\mathcal{D}_B$, as well as from $\mathcal{D}_B$ to $\mathcal{D}_A$, we can simply learn 2 sets of coupled mappings using MMCM, one for each scenario.

### B. Optimizing MMCM

We observe that the MMCM formulation shown in Eq. 6 is not convex. Because we are only guaranteed to find a local minimum when solving this non-convex optimization problem, we demonstrate the impact of different initializations for $\mathbf{P}_A$ and $\mathbf{P}_B$ in Sections VI and VII.

We find a local minimum using gradient descent. Algorithm 1 details the process of optimizing the MMCM approach. As mentioned earlier, we use the squared Euclidean distance in our experiments. With that distance metric in mind, the partial derivatives of the objective function $J$ with respect to $\mathbf{P}_A$ and $\mathbf{P}_B$ are

$$\begin{aligned}\nabla_{P_A} J = 2\lambda \sum_{M} (\mathbf{P}_A^T a_i - \mathbf{P}_B^T b_j) a_i^T \\ + 2(1-\lambda) \sum_{V_+} \mathbf{P}_B^T (b_k - b_j) a_i^T\end{aligned} \tag{9}$$

$$\begin{aligned}\nabla_{P_B} J = 2\lambda \sum_{M} (\mathbf{P}_B^T b_j - \mathbf{P}_A^T a_i) b_j^T \\ + 2(1-\lambda) \sum_{V_+} \Big( \mathbf{P}_B^T (b_j b_j^T - b_k b_k^T) \\ + \mathbf{P}_A^T a_i (b_k^T - b_j^T) \Big)\end{aligned} \tag{10}$$

where $V_+$ represents the subset of $V$ that has a non-zero hinge loss, based on $\mathbf{P}_{A,t}$ and $\mathbf{P}_{B,t}$, which are the projections at the $t^{\text{th}}$ iteration of the optimization. $\mathbf{P}_A$ and $\mathbf{P}_B$ are updated as follows,

$$\begin{aligned}\mathbf{P}_{A,t+1} = \mathbf{P}_{A,t} - \eta_t \nabla_{\mathbf{P}_A} J_t \\ \mathbf{P}_{B,t+1} = \mathbf{P}_{B,t} - \eta_t \nabla_{\mathbf{P}_B} J_t\end{aligned} \tag{11}$$

where $\eta$ is the learning rate. The learning rate is changed according to whether the update to $\mathbf{P}_A$ and $\mathbf{P}_B$ lowers the objective value; if the value is lowered, then $\eta_{t+1} = 1.1\eta_t$, and if the objective value is higher, we set $\eta_t = .9\eta_t$ and try updating $\mathbf{P}_{A,t+1}$ and $\mathbf{P}_{B,t+1}$ again. We stop the gradient descent algorithm when $\eta < \epsilon$, or after a determined number of iterations. In practice, we set $\epsilon = 10^{-6}$, and if that threshold is not reached, we stop after 150 iterations of the gradient descent algorithm.

*1) Approximated Gradient Descent:* One of the drawbacks with the gradient descent computation is that the size of $V_+$ is $O(n^3)$. With matrix multiplications performed for each member of $V_+$, this part of the gradient computation requires the most computation in the entire MMCM algorithm, and is a bottleneck for applying MMCM to larger datasets (which increases the size of $V_+$) or larger feature representations (increases the time each matrix multiplication takes). To avoid these problems, we instead choose to compute the gradients using only a subset of $V_+$. For each image $a_i \in A$, we use a fraction of $V_{i+}$, where $V_{i+} = \{(j,k) | y_{ai} = y_{bj}, y_{ai} \neq y_{bk}\}$ with a non-zero hinge loss. We denote the fraction used as $\rho = (0, 1]$, where $\rho = 1$ represents using the original gradient descent, $\rho = .5$ represents using half of the members of $V_{i+}$ for each image $a_i \in A$, and so on. Performing this approximated gradient descent can be effective when $V_+$ is very large, when even a small subset of $V_+$ can be representative of the true gradient. The risk of approximating

**Algorithm 1** Maximum-Margin Coupled Mapping

**Require:** $\mathbf{A}$ and $\mathbf{B}$ - data with labels $Y_A$ and $Y_B$, and $\lambda$ - trade-off parameter
Initialization: Need estimates of $\mathbf{P}_A$ and $\mathbf{P}_B$.
$\eta_0 = 1$
**while** (not converged) **do**
  $V_+ = \{(i,j,k) \in V, d(a_i, b_k) - d(a_i, b_j) < 1\}$
  **while** $J_{t+1} > J_t$ **do**
    $\mathbf{P}_{A,t+1} = \mathbf{P}_{A,t} - \eta_t \nabla_{\mathbf{P}_A} J_t$
    $\mathbf{P}_{B,t+1} = \mathbf{P}_{B,t} - \eta_t \nabla_{\mathbf{P}_B} J_t$
    **if** $J_{t+1} > J_t$ **then**
      $\eta_t = .9\eta_t$
    **end if**
  **end while**
  $\eta_{t+1} = 1.1\eta_t$
  $t = t + 1$
**end while**

gradient descent is present when the subset is too small to capture the true gradient. In this case, while the calculations in each iteration will be quick, the algorithm may not converge to a suitable result. We explored selecting the tuples from $V_{i+}$ based on the the largest hinge loss values, but found that picking a random sample was more effective. We discuss the effect of the $\rho$ parameter on both recognition performance and computation time in Section VII.

## V. Databases

We use four databases in this paper: two including only visible light images that we use to test matching faces at different resolutions, one comprised of visible light and near-infrared (NIR) photos, and one comprised of visible light photos and viewed sketches[1] Each test was run 10 times, with training images chosen randomly for each separate trial.

The general testing procedure is as follows. Images in both domains are run through the Viola-Jones face detector in MATLAB's Computer Vision System Toolbox [27]. The face detector outputs square bounding boxes. These regions are resized to a uniform size, and the left and right borders are equally cropped such that the remaining regions have an aspect ratio of 4:5. The borders are cropped because these areas often contained more background pixels than face information. Datasets containing larger images use principal component analysis (PCA) [6] to reduce the dimensionality of the data in both domains. When PCA is used, the basis vectors are learned on the training data for that trial, and we use enough basis vectors to keep 99% of the total variance. Each image is represented by either its PCA coefficients or pixel intensities. Once data is projected into the learned common subspace, matching is performed via nearest neighbor matching using the Euclidean distance in the common subspace as a measure of similarity between the probe and gallery set of images. The objective of our tests is to demonstrate which coupled mapping techniques learn the most effective mappings into a common subspace. More elaborate feature extraction techniques, such as that presented in [5], would almost certainly improve the performance of all the coupled mapping techniques presented.

We learn coupled mappings using an implementation of Canonical Correlation Analysis (CCA) by Sun et al. [28], as well as our own implementations of Simultaneous Discriminant Analysis (SDA) [21] and Coupled Marginal Fisher Analysis [22]. Each of these coupled mapping techniques are used as initializations for MMCM, and are denoted as $\text{MMCM}_{\text{CCA}}$ $\text{MMCM}_{\text{SDA}}$ and $\text{MMCM}_{\text{CMFA}}$ respectively.

For all tests using MMCM, we set $\lambda$ based on previous results showing that $\lambda = .5$ consistently performs very well [4]. We perform approximated gradient descent using $\rho = .05$, i.e., using 5% of triplets that violate the margin. The following sections will discuss the training, gallery, and probe sets for each dataset as well as other notes unique to each dataset.

We now briefly describe the different datasets we use to evaluate the efficacy of the proposed cross-domain image matching algorithm.

### A. Multi-PIE (Cross-Resolution)

The CMU Multi-PIE Face Database (Multi-PIE)[29] contains a total of 337 subjects captured over up to 4 sessions, with a number of different poses, illuminations, and facial expressions during each section. We use a subset of the Multi-PIE datasets to consider the problem of cross-resolution face recognition. Our training set is comprised of neutral expression photos from the first session, comprised of 249 subjects. We include all 20 illuminations, but only use 3 poses (yaw = $-15°$, $0°$, and $15°$). Because of the large number of images, face detections were not reviewed; instead any face detection that was smaller than 150 $\times$ 150 pixels was discarded (less than 1% of images). The remaining faces were normalized to 200 $\times$ 160 pixels, and 4 images were randomly selected from each subject to serve as training images and the test gallery. Low resolution training images were generated by downsampling and blurring the training images to 100 $\times$ 80 pixels. PCA is used to reduce the dimension of the data.

The training set for each trial is also used as the gallery set. We have two probe sets comprised of 100 $\times$ 80 pixel images from the second Multi-PIE session. 166 subjects appear in both the first and second sessions of the Multi-PIE dataset. Each of these subjects is imaged with the same 20 illuminations and 3 poses, as well as with 3 expressions, neutral, "surprise," and "squint". We report the results using neutral probe images separately from the other two expressions which are not represented in the training and gallery set.

### B. SCFace (Cross-Resolution)

The Surveillance Cameras Face Database (SCFace) [30] contains 130 subjects captured at low resolutions in a simulated setting with surveillance cameras. Each subject is captured at three distances from a portal, and 7 cameras (5 visible light, 2 near infrared (NIR)) positioned over the portal to photograph the subject. There are additional images for each subject, but we only use the 5 visible light surveillance cameras, and we designate the nearest distance as the high

[1] A viewed sketch is a sketch drawn by looking at an available photograph. (Compare to *forensic sketch*, where no photo is available.)

(a) Multi-PIE (b) SCFace (c) LDHF (d) CUFS

Fig. 3: Sample images from each database, after Viola-Jones face detection and cropping

resolution, and the farthest of distance as the low resolution. The high and low resolutions used are $90 \times 72$ and $30 \times 24$, respectively. We do not use PCA for this dataset. A total of 4 HR and 7 LR images did not have any faces detected. For these 11 images, bounding boxes are manually provided, and these are used in the gallery and training sets. These manual corrections are not used when the particular LR images are included in the probe sets.

We perform two tests with the SCFace database. In the first test, we select 4 sensors for all 130 subjects. The 520 HR and LR images comprise the training set, and the 520 HR images are used as the gallery set. The LR image from the remaining sensor for each of the 130 subjects (which varies for each subject) becomes the probe set. The second test uses 80 randomly selected subjects, and all 5 HR and LR images comprise the training set. From the remaining 50 subjects, a single HR images is chosen to be included in the gallery set, and 4 LR images from the remaining sensors are included in the probe set.

### C. LDHF Database (Visible to NIR)

The Long Distance Heterogeneous Face Database (LDHF) [31] is a dataset of 100 subjects captured at 4 distances with a visible light and NIR camera. We use high resolution images ($1280 \times 1024$) captured at 1 meter, and use PCA. A single visible light image did not have a face detected, and a manual bounding box was used instead. The training set is comprised of 67 subjects, and the 33 remaining visible light and NIR images make up the gallery and probe sets, respectively.

### D. CUFS (Photo to Sketch)

The CUHK Face Sketch database (CUFS) [32] contains 606 subjects with a single visible light images, and a corresponding viewed sketch. The resolution of both the photos and sketches is $320 \times 256$. There were no face detection errors for the 606 photos. Face detection was not performed on the sketches. Instead, we use the eye landmarks provided for both photos and sketches to align the sketches. After face detection was performed, we calculate the average eye location within the bounding boxes for the 606 photos. These average eye coordinates are used for all the sketches, which are then cropped and resized to the same resolution. The training set is comprised of 404 subjects, and the remaining 202 photos and sketches make up the gallery and probe sets, respectively.

## VI. Experimental Results

### A. Face Identification

Identification tasks arise when a person's biometric is captured, and we want to identify who the subject is among a closed set of people. In these scenarios, we can produce a rank ordering of gallery subjects, and determine the quality of an algorithm by where in the ordering the first true match is. We show the rank 1 identification rates for the SCFace "closed set" scenario, where all subjects are included in training, and Multi-PIE, in Table I, and cumulative match characteristic curves (CMCs) are shown in Figure 4. In addition to the coupled mappings, we also perform linear discriminant analysis (LDA) as a single-domain baseline, using only low-resolution images from the two databases for training and the gallery.

As Table I and Figure 4 show, MMCM improves on nearly every initialization in each of the three cases, with the exception of MDS in the SCFace database. As would be expected, probe images from Multi-PIE containing novel expressions were much more challenging than neutral expression probe images, which matched the neutral expression images used in training for Multi-PIE. The best results were produced by $\text{MMCM}_{\text{CCA}}$ for SCFace, and $\text{MMCM}_{\text{SDA}}$ for Multi-PIE.

Previous work [4] shows much lower rank 1 identification rates on the SCFace database, under a very similar testing regime. We attribute the large improvement to key factors. First, the images used in this work are higher resolution. Perhaps more significantly, the Viola-Jones face detector produces a wider and more consistent field of view (FOV). The wider FOV retains more information about each face; we found that reducing the FOV to only include a smaller face region reduced performance. Additionally, inspection of the manual landmarks in the SCFace show that even when eye coordinates are accurately annotated, they may not be reliable for aligning low resolution images, as a difference of 1 or 2 pixels can

drastically affect image appearance when the distance between the two eyes is normalized in alignment. This phenomenon can be seen in one of the subjects in Fig. 5. More extreme cases appear in the dataset but we do not have the permission to display those examples.

A separate note regarding LDA's CMC for SCFace (Fig. 4a): It can be seen that the identification rate only increases every $4^{\text{th}}$ rank. This is because the 4 gallery images for each subject are very tightly clustered after being projected into the LDA subspace, making it extremely likely that a probe image will have the 4 nearest neighbors belonging to the same subject, as well as the next 4 nearest neighbors, and so on.

| Method | SCFace | Multi-Pie, neutral | Multi-PIE, expressions |
|---|---|---|---|
| LDA | 85.37 | 61.46 | 36.67 |
| CCA | 84.30 | 9.23 | 6.43 |
| $\text{MMCM}_{\text{CCA}}$ | 95.19 | 68.77 | 42.03 |
| SDA | 91.61 | 69.45 | 42.94 |
| $\text{MMCM}_{\text{SDA}}$ | 91.16 | 74.17 | 45.06 |
| CMFA | 86.05 | 62.10 | 38.18 |
| $\text{MMCM}_{\text{CMFA}}$ | 92.91 | 68.94 | 40.69 |
| MDS | 49.27 | 59.13 | 35.12 |
| $\text{MMCM}_{\text{MDS}}$ | 47.67 | 64.22 | 36.87 |

TABLE I: Rank 1 identification rates (%) for each identification scenario.

| Method | SCFace | LDHF | CUFS |
|---|---|---|---|
| CCA | 36.17 | 21.21 | 25.69 |
| $\text{MMCM}_{\text{CCA}}$ | 27.48 | 16.76 | 23.91 |
| SDA | 30.57 | – | – |
| $\text{MMCM}_{\text{SDA}}$ | 31.32 | – | – |
| CMFA | 31.72 | 14.87 | 17.99 |
| $\text{MMCM}_{\text{CMFA}}$ | 28.45 | 15.06 | 17.87 |
| MDS | 35.50 | 15.44 | 12.92 |
| $\text{MMCM}_{\text{MDS}}$ | 36.36 | 15.53 | 12.92 |

TABLE II: Equal error rates (EER) for each verification scenario.

### B. Face Verification

Verification tasks differ from identification tasks in that, for verification the probe subject is not assumed to belong to a closed set of people. In this open set problem, we compare gallery images to probe images on a one-to-one basis, and use the match score to make a decision as to whether the two images are of the same person. Table II shows the equal error rates (EERs) for testing on the SCFace "open set" scenario, where 50 subjects are withheld from training, as well as the LDHF and CUFS datasets. Figure 6 shows the receiver operating characteristic curves (ROCs) for these tests. As a note, SDA is not trained on the LDHF and CUFS datasets, as these databases only contain a single image in each domain per subject.

While the results are not as consistently strong as the identification tests, $\text{MMCM}_{\text{CCA}}$ still achieves the lowest EER in the SCFace and CUFS datasets, while it is difficult to distinguish the best performance on the LDHF database between CMFA and $\text{MMCM}_{\text{CMFA}}$. The EERs are very high for both SCFace and LDHF, across all coupled mappings. This highlights how challenging these databases are, as designed for these experiments. The LDHF contains only 100 subjects, and using only 67 subjects for training makes learning a classifier that generalizes to unseen subjects very difficult. The poor performance of the coupled mappings on the SCFace database, when contrasted with 90% rank 1 identification rate in the "closed set" scenario highlights how the testing regime can dramatically impact results. The verification problem is made more challenging by (a) training on separate subjects from the testing set, and (b) only including 1 gallery image per test subject, and including only the 4 images from other cameras in the probe set. The low TAR rates at very low FARs could be attributed to high match scores between images from the same sensor, which can only occur between mismatched subjects.

| Method | Rank 1 Accuracy (%) | Equal Error Rate (%) |
|---|---|---|
| Random Projections | 0.49 (0.41) | 49.26 (1.16) |
| MMCM | 50.19 (3.21) | 11.08 (0.97) |

TABLE III: Results on CUFS sketch-to-photo matching using random coupled mappings as well as MMCM, using the random projections to initialize MMCM. We report the average ($\mu$ and $\sigma$) results over 100 different random sets.

## VII. Discussion

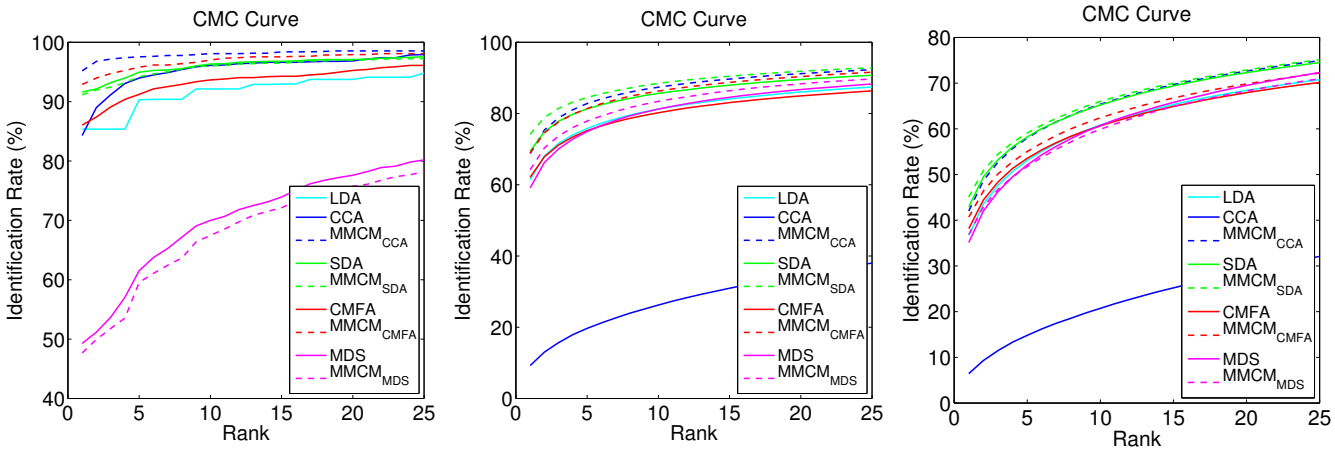
In this section, we will discuss a few additional properties of our proposed coupled mapping algorithm.

### A. Random Initializations

Due to the non-convex nature of the MMCM formulation, the MMCM optimization routine needs to be initialized by a "good" initial guess for the projections. Therefore in the previous section we used the projections learned by other coupled mapping methods as an initialization. In this subsection we investigate the sensitivity of our MMCM optimization routine to the initial projections that we start with by initializing the projections with random vectors. Table III shows the Rank-1 and EER performance of random coupled mappings and MMCM on the CUFS dataset. The results shown are obtained by using 100 different random sets of coupled mappings. While the random coupled mappings result in a performance that is no better than random guessing, MMCM initialized from these random projections is able to learn coupled mappings that perform well on the sketch-to-photo matching problem. Additionally, the low standard deviation in the performance of MMCM across the different random initializations suggests that MMCM is not very sensitive to the initializations. Further, the fact that the performance of MMCM when initialized with random projections is similar to the performance when initialized with projections from other coupled mapping methods suggests that even though the objective function is non-convex, there are many reasonably good solutions (local minima) that our optimization routine is able to converge to.

### B. Approximated Gradient Descent

We demonstrate the effect of the $\rho$ parameter on 5 trials on the SCFace database. SCFace is chosen because it is a larger




(a) SCFace "closed set" (b) Multi-PIE images, neutral (c) Multi-PIE images, expressions

Fig. 4: CMCs for each identification scenario

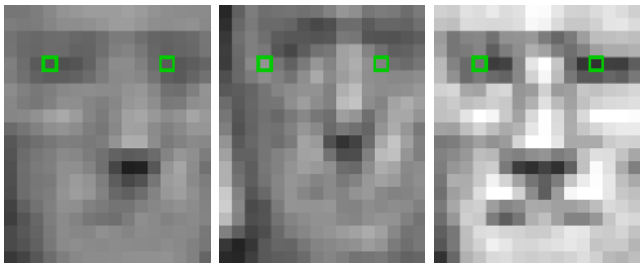


Fig. 5: Example of bad registration of SCFace images using manual landmarks, outlined in green. Small inconsistencies of 1 or 2 pixels result in very poorly aligned faces.

dataset, where $V_+$ can be very large and iterations of the full gradient descent can be time consuming. Figure 7 shows both the Rank 1 ID rates and the timing for optimizing MMCM. One plot shows that the Rank 1 ID rate is consistent, with a small dropoff from $\rho = 1$ to $\rho = .01$ ($< .5\%$). There is a sharper dropoff with very aggressive values of $\rho$ which result in the quickest optimizations. The timing measurements show that as the chosen value of $\rho$ is decreased, the time required consistently decreases. When $\rho = .05$, the value chosen for the tests in Section V, the time required is approximately half of the standard gradient descent. The speedup gained by low values of $\rho$ is reduced by the fact that some parts of the gradient computation are still independent of the value chosen, and computing the objective value uses the entire set of $V_+$ (necessary for choosing the correct value of $\eta_{t+1}$. Despite this, the choice of $\rho = .05$ results in a 2X speedup, with very little effect on the recognition performance.

## VIII. Conclusion

We discussed a new coupled mapping technique, MMCM, which is designed for cross-domain matching problems. We explained how the MMCM objective can be efficiently solved with an approximated gradient descent tailored to the challenges of the MMCM optimization. The effectiveness of MMCM for various cross-domain face biometric scenarios, including cross-resolution and cross-modality matching for both identification and verification problems. We also include face detection, making the results more indicative of an end-to-end biometrics system.

## References

[1] T. Wu, P. Turaga, and R. Chellappa, "Age estimation and face verification across aging using landmarks," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 6, pp. 1780–1788, 2012. 1

[2] U. Prabhu, J. Heo, and M. Savvides, "Unconstrained pose-invariant face recognition using 3d generic elastic models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 10, pp. 1952–1961, 2011. 1

[3] P. H. Hennings-Yeomans, S. Baker, and B. V. K. Vijaya Kumar, "Simultaneous super-resolution and feature extraction for recognition of low-resolution faces," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–8. 1, 2

[4] S. Siena, V. N. Boddeti, and B. V. Kumar, "Maximum-margin coupled mappings for cross-domain matching," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, September 2013. 1, 6, 7

[5] B. Klare and A. Jain, "Heterogenous face recognition using kernel prototype similarities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1410–1422, 2013. 1, 2, 6

[6] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991. 1, 6

[7] P. N. Belhumeur and J. Hespanha, "Eigenfaces vs. fisherfaces: Recognition using specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, July 1997. 1, 3

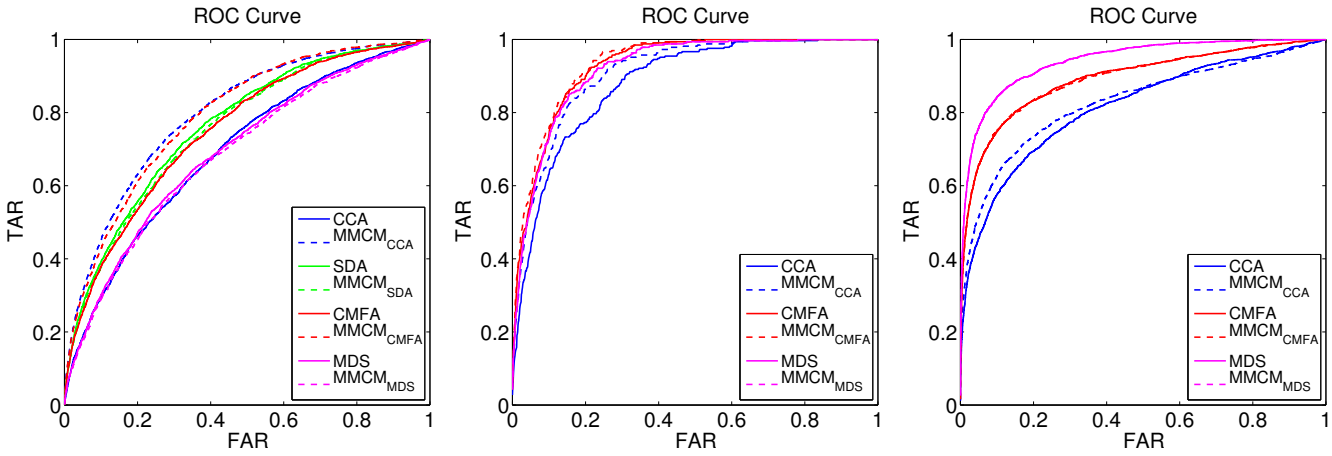
[8] K. Weinberger and L. Saul, "Distance metric learning for large margin nearest neighbor classification," *The Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009. 1, 4

[9] H. Daume III, "Frustratingly easy domain adaptation," in *45th Annual Meeting of the Association of Computational Linguistics*, June 2007, pp. 256–263. 2

[10] B. Kulis, K. Saenko, and T. Darrell, "What you saw is not what you get: Domain adaptation using asymmetric kernel transforms," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1785–1792. 2

[11] J. J. Lim, R. Salakhutdinov, and A. Torralba, "Transfer learning by borrowing examples for multiclass object detection," in *Neural Information Processing Systems (NIPS)*, 2011. 2

[12] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 689–696. 2




(a) SCFace "open set" (b) LDHF NIR to visible light (c) CUFS sketch to photo

Fig. 6: ROC curves for each verification scenario. Plots for $\text{MMCM}_{\text{MDS}}$ may be imposed over plots for MDS (see middle, right).

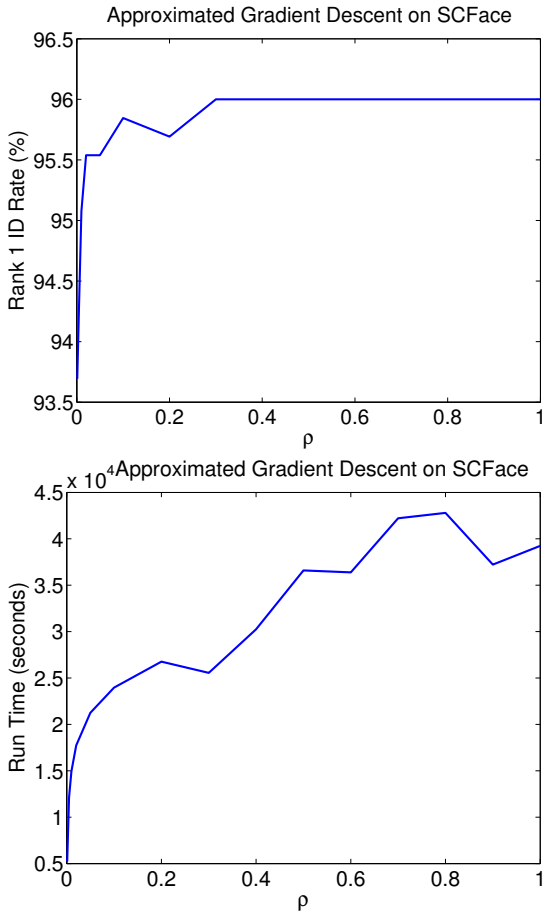



Fig. 7: Results on the SCFace "closed set" setting varying the $\rho$ parameter for approximated gradient descent.


[13] J. C. Pereira, E. Coviello, G. Doyle, N. Rasiwasia, G. R. Lanckriet, R. Levy, and N. Vasconcelos, "On the role of correlation and abstraction in cross-modal multimedia retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 521–535, 2014. 2

[14] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. 2

[15] J. Zuo, F. Nicolo, and N. Schmid, "Cross spectral iris matching based on predictive image mapping," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, 2010, pp. 1–5. 2

[16] J. Pillai, M. Puertas, and R. Chellappa, "Cross-sensor iris recognition through kernel learning," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 1, pp. 73–85, Jan 2014. 2

[17] H. Hotelling, "Relations between two sets of variants," *Biometrika*, vol. 328, no. 3, pp. 321–377, 1936. 2

[18] B. Li, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal Processing Letters*, vol. 17, no. 1, pp. 20–23, January 2010. 2

[19] Z. Yang, Y. Wu, Y. Jiang, Y. Zhou, L. Wang, W. Li, and Q. Liao, "Local consistency preserved coupled mappings for low-resolution face recognition," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, October 2013. 2

[20] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," in *European Conference on Computer Vision*, 2012. 2

[21] C. Zhou, Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Low-resolution face recognition via Simultaneous Discriminant Analysis," in *International Joint Conference on Biometrics*, 2011. 3, 6

[22] S. Siena, V. N. Boddeti, and B. V. Kumar, "Coupled marginal fisher analysis for low-resolution face recognition," in *European Conference on Computer Vision*, 2012. 3, 6

[23] S. Yan *et al.*, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, January 2007. 3

[24] S. Biswas, K. W. Bowyer, and P. J. Flynn, "Multidimensional scaling for matching low-resolution face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 2019–2030, 2012. 3

[25] I. Borg and P. J. F. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, second ed. ed. Springer, 2005. 3

[26] S. Parameswaran and K. Weinberger, "Large margin multi-task metric learning," in *Advances in Neural Information Processing Systems 23*, 2010, pp. 1867–1875. 4

[27] MATLAB and C. Toolbox, *version 8.1.0.604 (R2013a)*. Natick, Massachusetts: The MathWorks Inc., 2013. 6

[28] L. Sun, S. Ji, and J. Ye, "A least squares formulation for canonical correlation analysis," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1024–1031. 6

[29] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008. 6

[30] M. Grgic, K. Delac, and S. Grgic, "SCface - surveillance cameras face database," *Multimedia Tools and Applications Journal*, vol. 51, no. 3, pp. 863–879, February 2011. 6

[31] H. Maeng, S. Liao, D. Kang, S. Lee, and A. Jain, "Nighttime face recognition at long distance: Cross-distance and cross-spectral matching," in *Asian Conference on Computer Vision*, November 2012. 7

[32] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, 2009. 7