Maximum-Margin Coupled Mappings for Cross-Domain Matching

Stephen Siena, Vishnu Naresh Boddeti, B.V.K. Vijaya Kumar
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213 USA
ssiena@andrew.cmu.edu, naresh@cmu.edu, kumar@ece.cmu.edu

Abstract

Biometrics systems typically work best in settings where probe samples are captured in the same manner as the training set. When biometrics are acquired under different conditions or with different sensors, naïve approaches to recognition perform poorly. Coupled mappings have been introduced for performing face recognition across different resolutions, and learn a common subspace between different domains. In this paper, we introduce Maximum-Margin Coupled Mappings (MMCM), which aims to learn projections such that there is a margin of separation between pairs of cross-domain data from the same class and pairs of cross-domain data from different classes. While coupled mapping techniques have traditionally been used for matching face images at different resolutions, we demonstrate that MMCM is effective for cross-sensor biometric matching as well.

1. Introduction

A large focus of machine learning research is to perform matching between a probe sample and a gallery set of data for tasks such as information retrieval and recognition. Oftentimes the probe will lie in the same domain as the gallery set, but this is not always the case. This is often the case in many biometrics applications. For example, in some face recognition applications, the gallery images are of high resolution whereas the probe images may be captured at a lower resolution. When the probe is not of the same domain as the gallery, traditional classification methods trained on the gallery will not characterize the query in a way that allows for a useful comparison to the gallery set. In cases where the query has a different feature representation, it may be impossible to use methods otherwise available for tasks in a single domain.

The approaches for handling data from disparate domains can depend on the particular task or goal trying to be achieved. *Domain adaptation* and *transfer learning* refer to methods that attempt to learn a mapping from one domain to the other. In contrast, *coupled mappings* focus on learning a subspace that is common to both domains. 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 low-dimensional subspace that both domains can be mapped to.

This work introduces Maximum-Margin Coupled Mappings (MMCM), a new coupled mapping formulation inspired by large margin methods used in a single domain. The paper is structured as follows: Section 2 discusses similar work, particularly those related to coupled mappings. Section 3 provides the problem statement and how MMCM is designed to solve it. Section 4 shows empirical results on two biometrics tasks: low-resolution face recognition and cross-sensor ocular recognition. Section 5 concludes the paper.

2. Related Work

Domain adaptation and transfer learning methods have been used in a range of applications. It has grown in popularity in natural language processing [6], and has recently gained interest in computer vision tasks such as object detection and recognition [15, 17]. Pan & Yang provide a more thorough overview of the broader field [18].

There has also been work addressing cross-domain challenges in biometric applications. Early work by Hennings-Yeomans et al. performed super-resolution of faces for both reconstruction and recognition simultaneously [10]. Klare and Jain introduced a framework for heterogeneous face recognition using prototypes in multiple imaging modalities to compare different image representations [14]. Zuo et al. perform cross-spectral iris recognition by learning a predictive mapping function from the visible light spectrum to the near infrared channel [25].

Of particular interest to our work are other proposed methods of coupled mappings. As mentioned previously, coupled mappings aim to learn projections to a single subspace. Canonical Correlation Analysis (CCA) is a long-standing method that learns projections which maximize the correlation between two sets of data [11]. More recently,

Li et al. introduced two methods, one referred to as Coupled Mappings, and a more effective method called Coupled Locality Preserving Mappings (CLPM), which considered the local neighborhoods of data samples when learning projections [16]. Zhou et al. incorporated class labels into a coupled mapping objective function called Simultaneous Discriminant Analysis (SDA), and demonstrated a large improvement over CLPM [24]. Biswas et al. introduced a method based on multidimensional scaling that learns projections to a space such that distances between high and low-resolution face images approximate the distance between two corresponding high resolution images [2]. Finally, Kan et al. introduced Multi-view Discriminant Analysis (MvDA), a method to simultaneously learn sets of projections from any number of domains to a single common subspace [13].

With the exception of CCA, all the coupled mapping methods listed were introduced for face recognition. Much of the focus of coupled mapping is to leverage the information available in high-resolution face images to help recognize low-resolution face images. While this has traditionally been the goal of the previous coupled mapping methods, we intend to show the utility of these approaches in problems beyond face recognition.

Some coupled mapping techniques can be seen as extensions of single domain methods; for instance, the objective function of SDA is very similar to Linear Discriminant Analysis (LDA) [1]. Similarly, our work is inspired by Large Margin Nearest Neighbor (LMNN) classification [23]. The work learns a Mahalanobis distance such that same class pairs have a small distance while ensuring a margin between pairs of data from different classes. In Section 3, we discuss the relation of our work to LMNN.

3. Maximum-Margin Coupled Mappings

3.1. Problem Statement

In many scenarios, we have data samples 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 $\mathbf{B} = \{b_1, b_2, ..., b_n\} \in \mathcal{D}_B$, we want to be able to match samples from one domain to samples from the other. Defining a distance metric can be done simply enough when comparing data samples from the same domain, but it is often difficult to compare data across domains with different representations. To facilitate this, we can map data from both $\mathcal{D}_A$ and $\mathcal{D}_B$ into a common space, which we will denote $\mathcal{D}_Z \in \mathbb{R}^\gamma$. The goal is to then find functions $f_A : \mathcal{D}_A \rightarrow \mathcal{D}_Z$ and $f_B : \mathcal{D}_B \rightarrow \mathcal{D}_Z$. This paper introduces a method of learning linear projections, and 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 α-dimensional data sample a_i into a γ-dimensional subspace. Similarly, $\mathbf{P}_B \in \mathbb{R}^{\beta \times \gamma}$ projects the β-dimensional data sample b_i into a γ-dimensional subspace. It is not necessary that $\alpha = \beta$ nor that $\alpha \neq \beta$. 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 $\hat{\mathbf{B}} = \{\hat{b}_1, \hat{b}_2, ..., \hat{b}_m\} \in \mathcal{D}_Z$, data samples between the two sets can be naturally compared. Any range of classifiers can be trained using both sets of data, or simple metrics such as Euclidean distance 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 $Y_B = \{y_{b1}, y_{b2}, ..., y_{bn}\} \in \mathcal{L}$, we want to find projections such that in general $d_{y_{ai}=y_{bj}}(\hat{a}_i, \hat{b}_j) < d_{y_{ai}\neq y_{bk}}(\hat{a}_i, \hat{b}_k)$, where $d(\cdot, \cdot)$ is some distance metric. In other words, given a data sample from $\mathcal{D}_A$ and data samples from $\mathcal{D}_B$, we want the sample from $\mathcal{D}_A$ to be closer to those data samples from $\mathcal{D}_B$ that share the same class label.

3.2. MMCM Formulation

Canonical Correlation Analysis (CCA) is one of the oldest and most popular methods proposed to capture the correlations between data from different domains. However, because CCA is an unsupervised method, it cannot take advantage of class label information that might be available while training. Supervised coupled mapping methods like SDA [24], CMFA [20], and the proposed MMCM algorithm on the other hand can take advantage of the available class label information. SDA and CMFA are linear projection learning methods which try to minimize the distance between data samples with the same label and maximize the average distance between data samples with different labels, but are not designed to guarantee an explicit margin of separation between a match pair and an impostor pair. Inspired by the success of large margin based distance learning methods [19, 23], we propose to learn coupled mapping projections in a margin based formulation.

MMCM aims to find linear projections with the largest margin between match pairs across domains and non-match pairs across those domains. While other formulations for coupled mapping have drawn inspiration from traditional linear projection techniques such as Principal Component Analysis [22], MMCM draws upon support vector machines (SVMs) and other margin maximizing methods for its formulation. In learning projection matrices $\mathbf{P}_A$ and $\mathbf{P}_B$, we aim to solve the following minimization problem:

$$\min_{\mathbf{P}_A, \mathbf{P}_B} \lambda \frac{\sum_M d(a_i, b_j)}{|M|} + (1-\lambda) \frac{\sum_V [1 + d(a_i, b_j) - d(a_i, b_k)]_+}{|V|} \quad (1)$$

where $M = \{(i,j) | y_{ai} = y_{bj}\}$ and $V = \{(i,j,k) | y_{ai} = y_{bj}, y_{ai} \neq y_{bk}\}$ and $[x]_+$ represents the hinge loss, defined as $max(0, x)$. While any distance function can be used, we choose to use the squared Euclidean distance to compare

data samples, so we define

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

We can see in Eq. 1 that there are two terms: the first is a *pull* term that aims to project data samples from different domains close together when they share the same class label, and the second term is a *push* term that aims to keep projected data from different domains far apart when they do not share a class label. The second term is defined by both cross-domain match pairs and cross-domain impostor pairs. It penalizes cross-domain impostor pairs which lie closer to a_i than cross-domain matches by some margin. This margin is set to 1; it is well known in support vector machine (SVM) formulation that there is no loss of generality in setting the margin to 1. The *push* term uses a hinge loss; tuples in the set V that do not enter or cross that margin do not impact the computed objective value. λ is a parameter in the range $(0, 1)$ to determine the tradeoff between the two parts of the objective, and must be either set by a user or learned on a validation set.

3.3. Optimizing MMCM

It is important to first note that the objective function is not convex. Since solving the optimization problem can only guarantee finding a local minimum, it is important to initialize $\mathbf{P}_A$ and $\mathbf{P}_B$ with good estimates. This can include CCA or any other coupled mapping projections.

We choose to solve the optimization problem using gradient descent. Algorithm 1 gives an overview of the MMCM method. Keeping in mind that we use Euclidean distance in our optimization, the partial derivatives of the objective function J with respect to $\mathbf{P}_A$ and $\mathbf{P}_B$ are

$$\nabla_{P_A} J = 2\lambda \sum_M (\mathbf{P}_A^T a_i - \mathbf{P}_B^T b_j) a_i^T + 2(1-\lambda) \sum_{V_+} \mathbf{P}_B^T (b_k - b_j) a_i^T$$

$$\nabla_{P_B} J = 2\lambda \sum_M (\mathbf{P}_B^T b_j - \mathbf{P}_A^T a_i) b_j^T + 2(1-\lambda) \sum_{V_+} \mathbf{P}_B (b_j b_j^T - b_k b_k^T) + \mathbf{P}_A^T a_i (b_k^T - b_j^T)$$

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 learned projections at the t^{th} iteration. We update $\mathbf{P}_A$ and $\mathbf{P}_B$ as follows,

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

Algorithm 1 Maximum-Margin Coupled Mapping

Require: $\mathbf{A}$ and $\mathbf{B}$ - data with labels Y_A and Y_B, and λ - 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

where η is the learning rate. We control the learning rate according to whether each update to $\mathbf{P}_A$ and $\mathbf{P}_B$ lowers the objective value; if the value is lowered, then $\eta_{t+1} = 1.1\eta_t$, otherwise, we set $\eta_t = .9\eta_t$ and try updating $\mathbf{P}_{A,t+1}$ and $\mathbf{P}_{B,t+1}$ again. The gradient descent algorithm can continue indefinitely as $\eta \to 0$, so we halt the optimization after some number of iterations, or when $\eta < \epsilon$. Empirically, we found $\epsilon = 10^{-6}$ to be sufficient in our tests, with further iterations of the gradient descent algorithm having a very small impact on the projections in $\mathbf{P}_A$ and $\mathbf{P}_B$.

4. Experiments

In this section, we compare MMCM to other single-domain and cross-domain methods on two different tasks: cross-resolution face recognition and cross-sensor ocular recognition. We compare to LMNN using an implementation provided by the authors [23]. The other single-domain method used is LDA, a popular method for face recognition [1]. We compare to three cross-domain methods, CCA (implemented by Sun et al. [21]) and our own implementations of SDA [24] and CMFA [20]. Because the MMCM formulation is a non-convex optimization problem, it requires an initial estimate of the projections. In this work we initialize with projections learned by CCA, SDA, and CMFA. When MMCM_{CCA}, MMCM_{SDA}, or $\text{MMCM}_{\text{CMFA}}$ is written, it indicates that MMCM was initialized with projections learned with CCA, SDA, and CMFA, respectively.

4.1. Face Recognition

Low-resolution (LR) face recognition is an emerging research topic in biometrics. Methods for low-resolution face images have applications in long-distance recognition systems and systems that use low quality sensors, e.g. surveillance cameras. Methods that are successful in "ideal" face

Figure 1. Example of the wide variation in appearance of subjects in the Multi-PIE set used in Section 4.1 High resolution (HR) images are in the top row, with corresponding low resolution (LR) images in the bottom row.

recognition conditions fail at lower resolutions, due to the difficulty in getting texture information or accurate facial landmarks from such low-quality images. As an alternative to these standard approaches, many methods approach the problem by learning projections to a common subspace [2, 13, 20]. The assumption made in this approach is that when a LR query image is captured, the gallery will contain HR images. Under these conditions, the image can be matched by projecting the LR query image and the HR gallery to the same subspace.

4.1.1 Multi-PIE

We demonstrate the usefulness of MMCM for cross-resolution face recognition on the Multi-PIE (Pose, Illumination, and Expression) database [8]. The Multi-PIE database contains over 750,000 images of a total of 337 subjects, with a wide variety of appearance available for each subject. We use a subset of 158 subjects for which human-labeled eye centers used for image registration are available. The human-labelled landmarks allow faces to be cropped so images contain only the face, and allows the eyes to be in a fixed location. We also restrict the images to come from a single acquisition session. We use images from 3 poses (yaw = $-15°$, $0°$, and $15°$) and include 2 expressions, neutral and smiling. Images are captured with illuminations from 20 different directions, for a total of 120 images per subject, and a total of 18,960 images.

The training protocol in this work is to select a random subset of 4 images from each subject to use as training. HR images are resized to be 24×24.[1] LR images are generated by blurring and downsampling HR images to 12×12. Each method learns 143 projections using the 24×24 (HR) and 12×12 (LR) versions of the training images as cross-domain pairs. After projections are learned on the training set, the HR training set becomes a gallery for the LR query images, which are LR images created with the remaining 116 images (not used at all during the training) per subject. At test time, the nearest neighbor using L2 distance is found for each query image. The figure of merit used is the rank 1 identification rate, which is the percentage of query images that have a nearest neighbor with the same identity as the query image. Results are reported as the average of 10 trials, each using a random subset of the data as the training data and gallery.

Table 1 shows the rank 1 identification rates and Figure 2 shows the cumulative match characteristic curves (CMC), which represent how many probe images return a match within the first n matches. The 10 trials comprise over 180,000 tests of a single probe image, so the rank 1 identification rates have a standard deviation of approximately 0.11%.

The single-domain methods, LDA and LMNN, learn projections on LR images, and subsequently match the LR probe set to the LR gallery. While they perform better than the unsupervised CCA, they do not do as well as the supervised coupled mapping approaches. By leveraging class labels when learning projections, SDA and CMFA can more effectively learn the relation between images of different resolutions. SDA also outperforms LMNN and LDA, which cannot utilize the information available at the higher resolution.

MMCM is able to outperform all tested coupled mapping methods, with a rank 1 identification rate of over 80%. Unlike SDA and CMFA, which achieve similar results, MMCM aims to learn projections so that images from different classes are separated by a large margin, whereas SDA and CMFA optimize projections based on the average distance between pairs of data points.

The results also show that MMCM can improve on any of the tested coupled mapping methods for the application. The performance gain is most dramatic when applied to CCA - almost a 10% increase in the rank 1 identification rate. The performance gain is less pronounced when starting with SDA and CMFA, but this could be because these methods are already getting most of the valuable discriminative information from the data. The improvement for MMCM with any initialization is very large relative to the standard deviations of the rank 1 identification rates ($\sim .11\%$), so we can see that the improvement from MMCM is statistically significant with any initialization method.

[1] This is mainly for two reasons. First, Hennings-Yeomans et al. observed that rank-1 identification rates did not improve significantly for the Multi-PIE dataset when using images larger than 24×24 [10]. Second, this size keeps the computational complexity manageable while allowing for an investigation of various coupled mapping methods.

Table 1. Rank 1 identification rates on Multi-PIE face recognition task. Results are the averages of 10 trials. Δ indicates the improvement using MMCM versus the initial estimated projections.

Method	Rank-1 ID (%)	Δ
LMNN	70.49	–
LDA	71.33	–
CCA	63.67	–
MMCM_{CCA}	73.23	+9.56
SDA	76.41	–
MMCM_{SDA}	79.94	+3.53
CMFA	76.93	–
$\text{MMCM}_{\text{CMFA}}$	80.05	+3.12

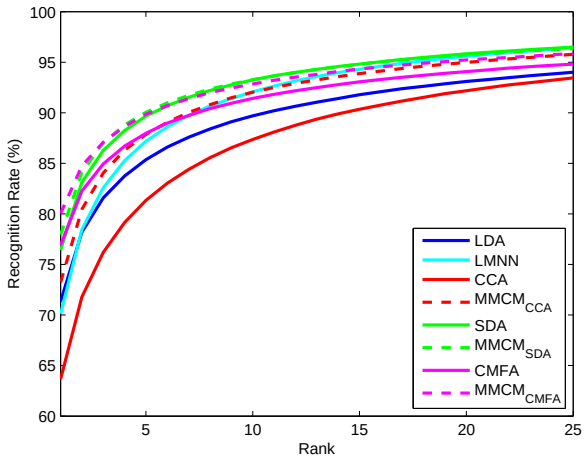

Figure 2. Cumulative match curves for Multi-PIE face recognition

4.1.2 SCface

One of the drawbacks of Multi-PIE is that there are no true LR images; the LR images are simply generated from a higher resolution. The SCface database [7] contains surveillance camera images of 130 subjects at different distances from the sensor, resulting in images of different resolutions. We use a subset of SCface consisting of 10 images per subject: 5 different visible light sensors capturing the subject at the second closest distance (HR), and the same sensors capturing the subject at the closest distance(LR). After cropping and aligning the images, the HR images are downsampled to 30×24, and the LR images are downsampled 15×12. Downsampling the images reduces overfitting of the projections due to the lack of training data, and a similar resolution to Multi-PIE allows for comparison of the two datasets. Figure 3 shows the 10 images for a single subject, and also illustrates the small differences in pose and appearance across the different sensors. Out-of-plane rotation of the face is also present, making face alignment less exact. All of these factors make the dataset very challenging.

We follow the same procedure as the Multi-PIE experiments. 4 sensors are chosen for each subject to be the training set, and the HR images become the gallery for matching the 130 probe images, which are comprised of the remaining sensor per subject. Results shown are with 129 projections. We chose parameters for both MMCM and the baselines that were learned on the Multi-PIE dataset.

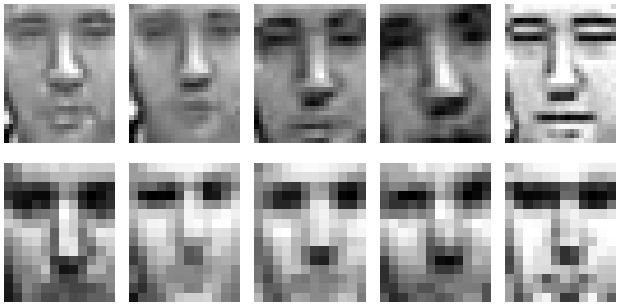

Figure 3. Images from SCface from a single subject. Each column is a different camera sensor. HR images are in top row and LR images are in the bottom row.

Table 2. Rank 1 identification rates on SCface face recognition task. Results are the averages of 10 trials. Δ indicates the improvement using MMCM versus the initial estimated projections.

Method	Rank-1 ID (%)	Δ
LMNN	31.62	–
LDA	22.85	–
CCA	19.77	–
MMCM_{CCA}	29.77	+10.00
SDA	35.08	–
MMCM_{SDA}	39.92	+4.84
CMFA	33.77	–
$\text{MMCM}_{\text{CMFA}}$	37.77	+4.00

Overall, the rank 1 rates are much lower compared to the Multi-PIE experiment due to the challenges posed in the SCface database. Still, there are similar trends as the Multi-PIE task, and MMCM outperforms the other coupled mapping baselines and the single-domain methods. Over 10 trials, there are 1,300 probe images tested, so the standard deviation of the rank 1 rates is approximately 1.39%, and the improvement seen with MMCM is once again statistically significant.

4.2. Ocular Recognition

Iris recognition is an increasingly popular technique for security applications. Iris cameras typically capture an infrared image of the eye and the surrounding region. After the image is captured, the iris is segmented and recognition can be performed [5]. While the accuracy of such methods can be very high, it has been shown that using the region surrounding the iris can produce higher recognition rates in challenging conditions [3]. Additionally, traditional iris recognition methods require high resolution im-

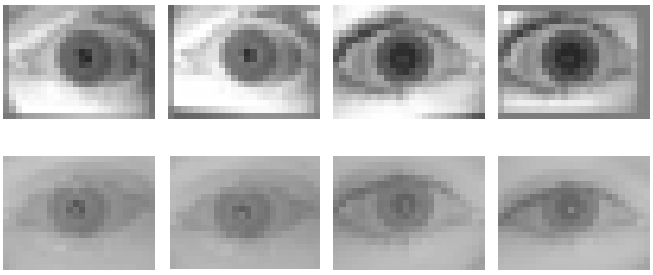

Figure 4. Example ocular images used for matching, after images are aligned according to eye corners. The top row shows images from the LG 4000 sensor, while the bottom row shows images from the CFAIRS sensor. The left 2 columns are authentic pairs of images from a single subject, and the right 2 columns are authentic pairs from a second subject.

ages [9]. The ocular region consists of the region surrounding the eye that is captured by an iris camera and captures approximately 15% of the face. We use ocular data from two sensors, the LG IrisAccess 4000 (LG 4000) [12] and the Honeywell Combined Face and Iris Recognition System (CFAIRS) [4], collected by the University of Notre Dame for the Biometrics Exploitation Science and Technology (BEST) Development Challenge Problem (BDCP). Sample images from each sensor can be seen in Figure 4. Both sensors collect images in the infrared spectrum, but there is a clear distinction in the appearance between images from each sensor. Additionally, images from the CFAIRS camera can be captured at a longer distance than the LG 4000. Images included in the test are not necessarily captured in a single session. The images are aligned according to the eye corners. Some images from the LG 4000 require some image padding after eye corner alignment, due to the tight field of view captured with the sensor. While the results shown will be including the padded images, tests using a smaller ocular region (and minimal image padding) showed slightly lower performance for all methods. There are a total of 2577 images from the CFAIRS sensor and 1559 images from the LG 4000.

The testing procedure is similar to that in Section 4.1. The dataset consists of 164 different ocular regions, and two LG 4000 images and two CFAIRS images are selected for each region to learn projections and also serve as the gallery sets. The other images become two different probe sets, one comprised of the remaining LG 4000 images, and one comprised of the remaining CFAIRS images. Once the projections are learned, two tests are performed: one matching the CFAIRS probe set to the LG 4000 gallery images, and one matching the LG 4000 probe set to the CFAIRS gallery images. As in the face recognition tests, this procedure is repeated 10 times with randomly selected gallery sets. Unlike face recognition across different resolutions, this cross-sensor matching includes two domains with the same dimensionality. This allows for training of single domain methods such as LMNN and LDA by simply pooling the LG 4000 and CFAIRS training images together for a total of 4 training images per ocular region. The matching protocol does not change; the only images included in the gallery are those not belonging to the sensor that produced the probe set.

Table 3. Rank 1 identification rates (in %) on ocular recognition task. The sensor name indicates the sensor of the probe set images.

Method	9×12		18×24	
	LG 4000	CFAIRS	LG 4000	CFAIRS
LDA	42.32	37.76	63.76	60.05
LMNN	54.53	54.23	57.77	57.41
CCA	75.91	49.42	80.80	75.60
MMCM_{CCA}	77.84	71.70	81.64	81.84
SDA	77.29	73.70	83.20	80.12
MMCM_{SDA}	**82.72**	**81.42**	84.76	82.35
CMFA	76.45	73.77	83.48	80.63
$\text{MMCM}_{\text{CMFA}}$	79.41	80.44	**85.05**	**82.92**

We compare MMCM's performance on cross-sensor ocular recognition at two different resolutions: learning 107 projections at 9×12 resolution, and learning 163 projections at 18×24 resolution. The rank 1 identification rates for all tests are shown in Table 3, and CMCs for the 9×12 resolution are shown in Figure 5. The standard deviations of the rank 1 identification rates are approximately 0.45% and 0.33% for the LG 4000 and CFAIRS probe sets, respectively. The trends for the ocular cross-sensor tests are similar to those seen in the cross-resolution face matching. Again the coupled mappings outperform the single-domain methods, and the supervised coupled mappings outperform CCA. MMCM once again achieves statistically significant improvements (>3 standard deviations) over its initializations. Also, not surprisingly, the 18×24 probe images yield higher recognition rates than 9×12 probe images.

4.3. Discussion

MMCM is a non-convex formulation for learning coupled mapping projections. In the two different tasks, we see that MMCM initialized with either CCA, SDA, or CMFA is able to find more discriminative projections. This is evident from the improved performance in those tasks. It is also important to note that MMCM will produce different projections depending on the initialization. The performance gains for each initialization is therefore derived from applying the large margin method in the local parameter space.

Another factor affecting the performance is the λ parameter weighing the *pull* and the *push* terms. As Figure 6 shows, the performance is not very sensitive to a particular value of λ, but instead degrades slowly from an optimal point. The consistent performance across the range of λ also

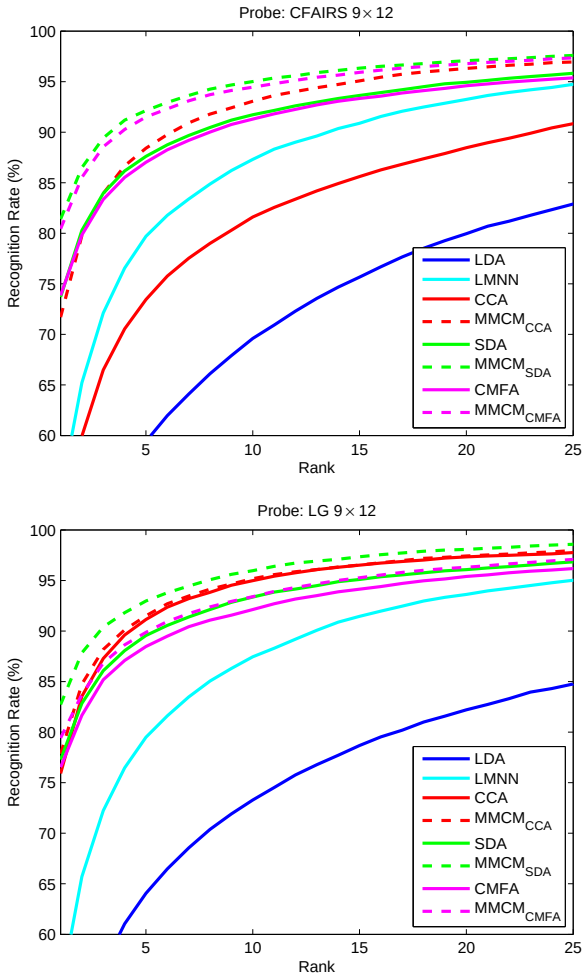

Figure 5. Cumulative match curves for ocular recognition

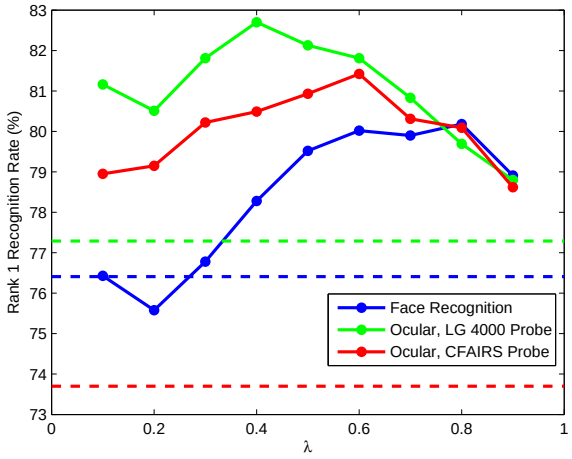

Figure 6. Performance on face and ocular recognition with varying λ. The solid line represents accuracy obtained with MMCM, while the dashed line represents the accuracy of the projections (learned via SDA) used to initialize MMCM. Ocular results are at 9×12 resolution.

demonstrates that both the *push* and *pull* terms of the optimization can contribute to the improved accuracy gained with MMCM.

5. Conclusions

This work has two primary contributions. First, it introduces a new method of learning coupled mappings for cross-domain matching, which aims to maximize the margin between match pairs and impostor pairs. Second, it demonstrates that coupled mappings can be used on tasks beyond low resolution face recognition, typically the focus of such methods. The effectiveness of MMCM is demonstrated on two different tasks and shows that it can learn more discriminative projections compared to other coupled mapping projection learning algorithms.

References

[1] P. N. Belhumeur and J. Hespanha. Eigenfaces vs. fisherfaces: Recognition using specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997. 2, 3

[2] 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*, 34(10):2019–2030, 2012. 2, 4

[3] V. Boddeti, J. Smereka, and B. V. K. Vijaya Kumar. A comparative evaluation of iris and ocular recognition methods on challenging ocular images. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8, oct. 2011. 5

[4] C. Boehnen, D. Barstow, D. Patlolla, and C. Mann. A multi-sample standoff multimodal biometric system. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pages 127–134, 2012. 6

[5] J. Daugman. How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):21–30, 2004. 5

[6] H. Daume III. Frustratingly easy domain adaptation. In *45th Annual Meeting of the Association of Computational Linguistics*, pages 256–263, June 2007. 1

[7] M. Grgic, K. Delac, and S. Grgic. SCface - surveillance cameras face database. *Multimedia Tools and Applications Journal*, 51(3):863–879, February 2011. 5

[8] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008. 4

[9] P. Grother, E. Tabassi, G. Quinn, and W. Salamon. Performance of iris recognition algorithms on standard images. September 2009. Technical report, published as NIST Interagency Report 7629. 6

[10] 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*, pages 1–8, 2008. 1, 4

[11] H. Hotelling. Relations between two sets of variants. *Biometrika*, 328(3):321–377, 1936. 1

[12] Iris ID Inc. Irisaccess 4000, February 2012. 6

[13] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen. Multi-view discriminant analysis. In *European Conference on Computer Vision*, 2012. 2, 4

[14] B. Klare and A. Jain. Heterogenous face recognition using kernel prototype similarities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1410–1422, 2013. 1

[15] 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*, pages 1785–1792, 2011. 1

[16] B. Li, S. Shan, and X. Chen. Low-resolution face recognition via coupled locality preserving mappings. *IEEE Signal Processing Letters*, 17(1):20–23, January 2010. 2

[17] J. J. Lim, R. Salakhutdinov, and A. Torralba. Transfer learning by borrowing examples for multiclass object detection. In *Neural Information Processing Systems (NIPS)*, 2011. 1

[18] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010. 1

[19] S. Parameswaran and K. Weinberger. Large margin multi-task metric learning. In *Advances in Neural Information Processing Systems 23*, pages 1867–1875. 2010. 2

[20] S. Siena, V. N. Boddeti, and B. V. K. Vijaya Kumar. Coupled marginal fisher analysis for low-resolution face recognition. In *European Conference on Computer Vision*, 2012. 2, 3, 4

[21] 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*, pages 1024–1031. ACM, 2008. 3

[22] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991. 2

[23] K. Weinberger and L. Saul. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research*, 10:207–244, 2009. 2, 3

[24] 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. 2, 3

[25] 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*, pages 1–5, 2010. 1