Dictionary-based Domain Adaptation Methods for the Re-identification of Faces

Qiang Qiu, Jie Ni, and Rama Chellappa

Abstract Re-identification refers to the problem of recognizing a person at a different location after one has been captured by a camera at a previous location. We discuss re-identification of faces using the domain adaptation approach which tackles the problem where data in the target domain (different location) are drawn from a different distribution as the source domain (previous location), due to different view points, illumination conditions, resolutions, etc. In particular, we discuss the adaptation of dictionary-based methods for re-identification of faces. We first present a domain adaptive dictionary learning (DADL) framework for the task of transforming a dictionary learned from one visual domain to the other, while maintaining a domain-invariant sparse representation of a signal. Domain dictionaries are modeled by a linear or non-linear parametric function. The dictionary function parameters and domain-invariant sparse codes are then jointly learned by solving an optimization problem. We then discuss an unsupervised domain adaptive dictionary learning (UDADL) method where labeled data are only available in the source domain. We propose to interpolate subspaces through dictionary learning to link the source and target domains. These subspaces are able to capture the intrinsic domain shift and form a shared feature representation for cross domain identification.

Qiang Qiu
the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708. e-mail: qiang.qiu@duke.edu

Jie Ni and Rama Chellappa
the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742. They are also with the Center for Automation Research, UMIACS, University of Maryland, College Park, MD 20742 e-mail: {jni, rama}@umiacs.umd.edu

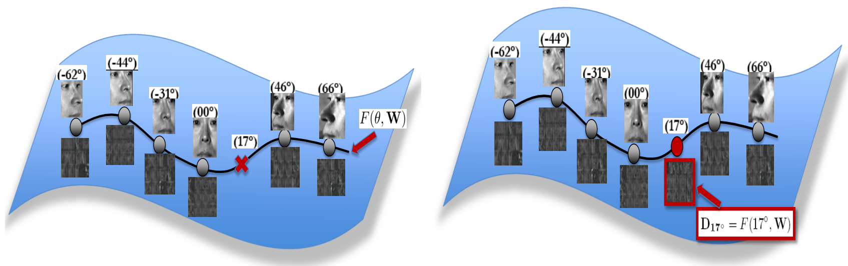
1 Introduction

Re-identification refers to identify a subject initialized at one location with a feasible set of candidates at other locations and over time. We are interested in face re-identification as face is an important biometric signature to determine the identity of a person. Re-identification is a fundamentally challenging problem due to the large visual appearance changes caused by variations in view angle, lighting, background clutter and occlusion [37]. It is well known that traditional face recognition techniques perform well when constrained face images are acquired at close range, with controlled studio lights and cooperative subjects. Yet these ideal assumptions are usually violated in the scenario of re-identification, which poses serious challenges to standard face recognition algorithms [5]. As it is very difficult to address the large appearance changes through physical models of individual degradations, we formulate the face re-identification problem as a domain adaptation problem to handle the distribution shift between query and candidate images.

Domain Adaptation (DA) aims to utilize a *source domain* (early location) with plenty of labeled data to learn a classifier for a *target domain* (different location) which belongs to a different distribution. It has drawn much attention in the computer vision community [28, 16, 13, 12]. Based on the availability of labeled data in the target domain, DA methods can be classified into two categories: *semi-supervised* and *unsupervised* DA. Semi-supervised DA leverages the few labels in the target data or correspondence between the source and target data to reduce the divergence between two domains. Unsupervised DA is inherently a more challenging problem without any labeled target data to build associations between two domains.

In this chapter, we investigate the DA problem using dictionary learning and sparse representation approaches. Sparse and redundant modeling of signals has received a lot of attention from the vision community [33]. This is mainly due to the fact that signals or images of interest are sparse or compressible in some dictionary. In other words, they can be well approximated by a linear combination of a few atoms of a redundant dictionary. It has been observed that dictionaries learned directly from data achieved state-of-the-art results in a variety of tasks in image restoration [19, 9] and classification [34, 36].

When designing dictionaries for image classification tasks, we are often confronted with situations where conditions in the training set are different from those present during testing. For example, in the case of face re-identification, more than one familiar view may be available for training. Such training faces may be obtained from a live or recorded video sequences, where a range of views are observed. However, the test images can contain conditions that are not necessarily presented in the training images such as a face in a different pose. For such cases where the same set of signals are observed in several visual domains with correspondence information available, we discuss the proposed domain adaptive dictionary learning (DADL) method in [26] to learn a dictionary for a new domain associated with no observations. We formulate this problem of dictionary transformation in a function learning framework, i.e., dictionaries across different domains are modeled by a parametric function. The dictionary function parameters and domain-invariant sparse codes



(a) Example dictionaries learned at known poses with observations.

(b) Domain adapted dictionary at a pose ($\theta = 17^\circ$) associated with no observations.

Fig. 1: Overview of DADL. Consider example dictionaries corresponding to faces at different azimuths. (a) shows a depiction of example dictionaries over a curve on a dictionary manifold which will be discussed later. Given example dictionaries, our approach learns the underlying dictionary function $F(\theta, \mathbf{W})$. In (b), the dictionary corresponding to a domain associated with observations is obtained by evaluating the learned dictionary function at the corresponding domain parameters [26].

are then jointly learned by solving an optimization problem. As shown in Figure 1, given a learned dictionary function, a dictionary adapted to a new domain is obtained by evaluating such a dictionary function at the corresponding domain parameters, e.g., pose angles. The domain invariant sparse representations are used here as shared feature representation for cross domain face re-identification.

We further discuss the unsupervised DA with no correspondence information or labeled data in the target domain. Unsupervised DA is more representative of real-world scenarios for re-identification. In addition to individual degradation factors due to view points, lighting, resolution etc, sometimes the coupling effect among these different factors give rise to more variations in the target domain. As it is very costly to obtain labels for target images under all kinds of acquisition condition, it is more desirable that our identification system can adapt in an unsupervised fashion. We discuss an unsupervised domain adaptive dictionary learning (UDADL) method to learn a set of intermediate domain dictionaries between the source and target domains, as illustrated in Figure 2. We then apply invariant sparse codes across the source, intermediate and target domains to render intermediate representations, which provide a shared feature space for face re-identification. A more detailed discussion of UDADL can be found in [20].

1.1 Sparse Representation

Sparse signal representations have recently drawn much attention in vision, signal and image processing [1], [27], [25], [33]. This is mainly due to the fact that signals and images of interest can be sparse in some dictionary. Given an over-complete

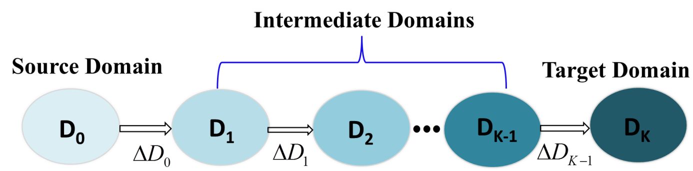


Fig. 2: Given labeled data in the source domain and unlabeled data in the target domain, our DA procedure learns a set of intermediate domains (represented by dictionaries $\{\mathbf{D}_k\}_{k=1}^{K-1}$) and the target domain (represented by dictionary $\mathbf{D}_K$) to capture the intrinsic domain shift between two domains. $\{\Delta\mathbf{D}_k\}_{k=0}^{K-1}$ characterize the gradual transition between these subspaces.

dictionary $\mathbf{D}$ and a signal $\mathbf{y}$, finding a sparse representation of $\mathbf{y}$ in $\mathbf{D}$ entails solving the following optimization problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \mathbf{D}\mathbf{x}, \tag{1}$$

where the ℓ_0 sparsity measure $\|\mathbf{x}\|_0$ counts the number of nonzero elements in the vector $\mathbf{x}$. Problem (1) is NP-hard and cannot be solved in a polynomial time. Hence, approximate solutions are usually sought [1], [6], [24], [30].

The dictionary $\mathbf{D}$ can be either based on a mathematical model of the data [1] or it can be trained directly from the data [21]. It has been observed that learning a dictionary directly from training rather than using a predetermined dictionary (such as wavelet or Gabor) usually leads to better representation and hence can provide improved results in many practical applications such as restoration and classification [27], [33].

Various algorithms have been developed for the task of training a dictionary from examples. One of the most commonly used algorithms is the K-SVD algorithm [1]. Let $\mathbf{Y}$ be a set of N input signals in a n-dimensional feature space $\mathbf{Y} = [\mathbf{y_1}...\mathbf{y_N}],\ \mathbf{y_i} \in \mathbb{R}^n$. In K-SVD, a dictionary with a fixed number of K items is learned by finding a solution iteratively to the following problem:

$$\arg\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_\mathbf{F}^\mathbf{2} \quad s.t.\ \forall i, \|\mathbf{x_i}\|_0 \le T \tag{2}$$

where $\mathbf{D} = [\mathbf{d_1}...\mathbf{d_K}],\ \mathbf{d_i} \in \mathbb{R}^n$ is the learned dictionary, $\mathbf{X} = [\mathbf{x_1}, ..., \mathbf{x_N}],\ \mathbf{x_i} \in \mathbb{R}^K$ are the sparse codes of input signals $\mathbf{Y}$, and T specifies the sparsity that each signal has fewer than T items in its decomposition. Each dictionary atom $\mathbf{d_i}$ is l_2-normalized.

Organization of the chapter: The structure of the rest of the chapter is as follows: in Section 2, we relate our work to existing work on domain adaptation. In Section 3, we discuss the domain adaptive dictionary learning framework for domain adaptation with correspondence available. In Section 4, we present the details of our unsupervised domain adaptive dictionary learning method. We report experimental results on face pose alignment and face re-identification in Section 5. The chapter is summarized in Section 6.

2 Related work

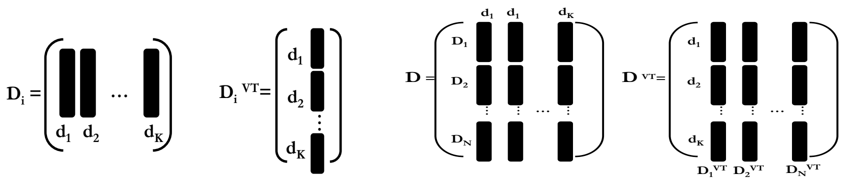
Several DA methods have been discussed in the literature. We briefly review relevant work below. Semi-supervised DA methods rely on labeled target data or correspondence between two domains to perform cross domain classification. Daume [7] proposes a feature augmentation technique such that data points from the same domain are more similar than those from different domains. The Adaptive-SVM introduced in [35] selects the most effective auxiliary classifiers to adapt to the target dataset. The method in [8] designed an adaptive classifier based on multiple base kernels. Metric learning approaches were also proposed [28, 16] to learn a cross domain transformation to link two domains. Recently, Jhuo et al. [15] utilized low rank reconstructions to learn a transformation so that the transformed source samples can be linearly reconstructed by the target samples.

Given no labels in the target domain to learn the similarity measure between data instances across domains, unsupervised DA is more difficult to tackle. It usually enforces certain prior assumption to relate the source and target data. Structural correspondence learning [4] induces correspondence among features from the two domains by modeling their relations with *pivot* features, which appear frequently in both domains. Manifold-alignment based DA [32] computes similarity between data points in different domains through the local geometry of data points within each domain. The techniques in [22, 23] learn a latent feature space where domain similarity is measured using maximum mean discrepancy. Two recent approaches [13], [12] in the computer vision community are more relevant to our methodology of UDADL, where the source and target domains are linked by sampling finite or infinite number of intermediate subspaces on the Grassmannian manifold. These intermediate subspaces are able to capture the intrinsic domain shift. Compared to their abstract manifold walking strategies, our UDADL approach emphasizes on synthesizing intermediate subspaces in a manner which gradually reduces the reconstruction error of the target data.

3 Domain Adaptive Dictionary Learning

We denote the same set of P signals observed in N different domains as $\{\mathbf{Y_1}, ..., \mathbf{Y_N}\}$, where $\mathbf{Y_i} = [\mathbf{y_{i1}}, ..., \mathbf{y_{iP}}]$, $\mathbf{y_{ip}} \in \mathbb{R}^n$. Thus, $\mathbf{y_{ip}}$ denotes the p^{th} signal observed in the i^{th} domain. In the following, we will use $\mathbf{D_i}$ as the vector-space embedded dictionary. Let $\mathbf{D_i}$ denote the dictionary for the i^{th} domain, where $\mathbf{D_i} = [\mathbf{d_{i1}} ... \mathbf{d_{iK}}]$, $\mathbf{d_{ik}} \in \mathbb{R}^n$. We define a *vector transpose* (VT) operation over dictionaries as illustrated in Figure 3. The VT operator treats each individual dictionary atom as a value and then perform the typical matrix transpose operation. Let $\mathbf{D}$ denote the stack dictionary shown in Figure 3b over all N domains. It is noted that $\mathbf{D} = [\mathbf{D^{VT}}]^{\mathbf{VT}}$.

The domain dictionary learning problem can be formulated as (3). Let $\mathbf{X} = [\mathbf{x_1}, ..., \mathbf{x_P}]$, $\mathbf{x_p} \in \mathbb{R}^K$, be the sparse code matrix. The set of domain dictionary



(a) $\mathbf{D_i}$ vs. $\mathbf{D_i}^{VT}$ **(b)** $\mathbf{D}$ vs. $\mathbf{D}^{VT}$

Fig. 3: The vector transpose (VT) operator over dictionaries.

$\{\mathbf{D_i}\}_i^N$ learned through (3) enable the same sparse codes $\mathbf{x_p}$ for a signal $\mathbf{y_p}$ observed across N different domains to achieve domain adaptation.

$$\arg\min_{\{\mathbf{D_i}\}_\mathbf{i}^\mathbf{N}, \mathbf{X}} \sum_i^N \|\mathbf{Y_i} - \mathbf{D_i X}\|_F^2 \;\; s.t. \; \forall p \; \|\mathbf{x}_p\|_o \leq T, \tag{3}$$

where $\|\mathbf{x}\|_o$ counts the number of non-zero values in $\mathbf{x}$. T is a sparsity constant.

We propose to model domain dictionaries $\mathbf{D_i}$ through a parametric function in (4), where $\boldsymbol{\theta_i}$ denotes a vector of domain parameters, e.g., view point angles, illumination conditions, etc., and $\mathbf{W}$ denotes the dictionary function parameters.

$$\mathbf{D_i} = F(\boldsymbol{\theta_i}, \mathbf{W}) \tag{4}$$

Applying (4) to (3), we formulate the domain dictionary function learning as (5).

$$\arg\min_{\mathbf{W}, \mathbf{X}} \sum_i^N \|\mathbf{Y_i} - F(\boldsymbol{\theta_i}, \mathbf{W})\mathbf{X}\|_F^2 \;\; s.t. \; \forall p \; \|\mathbf{x}_p\|_o \leq T. \tag{5}$$

Once a dictionary is estimated it is projected back to the dictionary-space by the projection operation described earlier.

We adopt power polynomials to model $\mathbf{D_i^{VT}}$ in Figure 3a through the following dictionary function $F(\boldsymbol{\theta_i}, \mathbf{W})$,

$$F(\boldsymbol{\theta_i}, \mathbf{W}) = w_0 + \sum_{s=1}^S w_{1s}\theta_{is} + ... + \sum_{s=1}^S w_{ms}\theta_{is}^m \tag{6}$$

where we assume S-dimensional domain parameter vectors and an m^{th}-degree polynomial model. For example, given $\boldsymbol{\theta_i}$ a 2-dimensional domain parameter vector, a quadratic dictionary function is defined as,

$$F(\boldsymbol{\theta_i}, \mathbf{W}) = w_0 + w_{11}\theta_{i1} + w_{12}\theta_{i2} + w_{21}\theta_{i1}^2 + w_{22}\theta_{i2}^2$$

Given $\mathbf{D_i}$ contains K atoms and each dictionary atom is in the $\mathbb{R}^n$ space, as $\mathbf{D_i^{VT}} = F(\boldsymbol{\theta_i}, \mathbf{W})$, it can be noted from Figure 3 that w_{ms} is a nK-sized vector. We define the function parameter matrix $\mathbf{W}$ and the domain parameter matrix $\boldsymbol{\Theta}$ as

$$\mathbf{W} = \begin{bmatrix} w_0^{(1)} & w_0^{(2)} & w_0^{(3)} & ... & w_0^{(nK)} \\ w_{11}^{(1)} & w_{11}^{(2)} & w_{11}^{(3)} & ... & w_{11}^{(nK)} \\ & & & & . \\ & & & & . \\ & & & & . \\ w_{mS}^{(1)} & w_{mS}^{(2)} & w_{mS}^{(3)} & ... & w_{mS}^{(nK)} \end{bmatrix} \qquad \boldsymbol{\Theta} = \begin{bmatrix} 1 & 1 & 1 & ... & 1 \\ \theta_{11} & \theta_{21} & \theta_{31} & ... & \theta_{N1} \\ & & & & . \\ & & & & . \\ & & & & . \\ \theta_{1S}^m & \theta_{2S}^m & \theta_{3S}^m & ... & \theta_{NS}^m \end{bmatrix}$$

Each row of $\mathbf{W}$ corresponds to the nK-sized w_{ms}^T, and $\mathbf{W} \in \mathbb{R}^{(mS+1)\times nK}$. N different domains are assumed and $\boldsymbol{\Theta} \in \mathbb{R}^{(mS+1)\times N}$. With the matrix $\mathbf{W}$ and $\boldsymbol{\Theta}$, (6) can be written as,

$$\mathbf{D^{VT}} = \mathbf{W^T}\boldsymbol{\Theta} \tag{7}$$

where $\mathbf{D^{VT}}$ is defined in Figure 3b. Now dictionary function learning formulated in (5) can be written as,

$$\arg\min_{\mathbf{W},\mathbf{X}} \|\mathbf{Y} - [\mathbf{W^T}\boldsymbol{\Theta}]^{\mathbf{VT}}\mathbf{X}\|_F^2 \;\; s.t. \, \forall p \; \|\mathbf{x}_p\|_o \leq T \tag{8}$$

where $\mathbf{Y}$ is the stacked training signals observed in different domains. With the objective function defined in (8), the dictionary function learning can be performed as described below.

Step 1: Obtain the sparse coefficients X and $[\mathbf{W^T}\boldsymbol{\Theta}]^{\mathbf{VT}}$ via any dictionary learning method, e.g., K-SVD [1].

Step 2: Given the domain parameter matrix $\boldsymbol{\Theta}$, the optimal dictionary function can be obtained as [18],

$$\mathbf{W} = [\boldsymbol{\Theta}\boldsymbol{\Theta}^{\mathbf{T}}]^{-\mathbf{1}}\boldsymbol{\Theta}[[[\mathbf{W^T}\boldsymbol{\Theta}]^{\mathbf{VT}}]^{\mathbf{VT}}]^{\mathbf{T}}. \tag{9}$$

4 Unsupervised Domain Adaptive Dictionary Learning

In this section, we present the UDADL method for face re-identification. We first describe some notations to facilitate subsequent discussions.

Let $\mathbf{Y}_s \in \mathbb{R}^{n*N_s}$, $\mathbf{Y}_t \in \mathbb{R}^{n*N_t}$ be the data instances from the source and target domain respectively, where n is the dimension of the data instance, N_s and N_t denote the number of samples in the source and target domains. Let $\mathbf{D}_0 \in \mathbb{R}^{n*m}$ be the dictionary learned from $\mathbf{Y}_s$ using standard dictionary learning methods, e.g, K-SVD [1], where m denotes the number of atoms in the dictionary.

We hypothesize there is a virtual path which smoothly connects the source and target domain. Imagine the source domain consists of face images in the frontal view while the target domain contains those in the profile view. Intuitively, face images which gradually transform from the frontal to profile view will form a smooth transition path. Our approach samples several intermediate domains along this virtual path, and associate each intermediate domain with a dictionary $\mathbf{D}_k, k \in [1, K]$, where K is the number of intermediate domains.

4.1 Learning Intermediate Domain Dictionaries

Starting from the source domain dictionary $\mathbf{D}_0$, we learn the intermediate domain dictionaries $\{\mathbf{D}_k\}_{k=1}^K$ sequentially to gradually adapt to the target data. This is also conceptually similar to incremental learning. The final dictionary $\mathbf{D}_K$ which best represents the target data in terms of reconstruction error is taken as the target domain dictionary. Given the k-th domain dictionary $\mathbf{D}_k, k \in [0, K-1]$, we learn the next domain dictionary $\mathbf{D}_{k+1}$ based on its coherence with $\mathbf{D}_k$ and the remaining residue of the target data. Specifically, we decompose the target data $\mathbf{Y}_t$ with $\mathbf{D}_k$ and get the reconstruction residue $\mathbf{J}_k$:

$$\mathbf{\Gamma}_k = \arg\min_{\mathbf{\Gamma}} \|\mathbf{Y}_t - \mathbf{D}_k\mathbf{\Gamma}\|_F^2, s.t. \forall i, \|\alpha_i\|_0 \leq T \tag{10}$$

$$\mathbf{J}_k = \|\mathbf{Y}_t - \mathbf{D}_k\mathbf{\Gamma}_k\|_F^2 \tag{11}$$

where $\mathbf{\Gamma}_k = [\alpha_1, ..., \alpha_{N_t}] \in \mathbb{R}^{m*N_t}$ denote the sparse coefficients of $\mathbf{Y}_t$ decomposed with $\mathbf{D}_k$, and T is the sparsity level. We then obtain $\mathbf{D}_{k+1}$ by estimating $\varDelta\mathbf{D}_k$, which is the adjustment in the dictionary atoms between $\mathbf{D}_{k+1}$ and $\mathbf{D}_k$:

$$\min_{\varDelta\mathbf{D}_k} \|\mathbf{J}_k - \varDelta\mathbf{D}_k\mathbf{\Gamma}_k\|_F^2 + \lambda\|\varDelta\mathbf{D}_k\|_F^2 \tag{12}$$

This formulation consists of two terms. The first term ensures that the adjustments in the atoms of $\mathbf{D}_k$ will further decrease the current reconstruction error $\mathbf{J}_k$. The second term penalizes abrupt changes between adjacent intermediate domains, so as to obtain a smooth path. The parameter λ controls the balance between these two terms. This is a ridge regression problem. By setting the first order derivatives to be zeros, we obtain the following closed form solution:

$$\varDelta\mathbf{D}_k = \mathbf{J}_k\mathbf{\Gamma}_k^T(\lambda\mathbf{I} + \mathbf{\Gamma}_k\mathbf{\Gamma}_k^T)^{-1} \tag{13}$$

where $\mathbf{I}$ is the identity matrix. The next intermediate domain dictionary $\mathbf{D}_{k+1}$ is then obtained as:

$$\mathbf{D}_{k+1} = \mathbf{D}_k + \varDelta\mathbf{D}_k \tag{14}$$

Starting from the source domain dictionary $\mathbf{D}_0$, we apply the above adaptation framework iteratively, and terminate the procedure when the magnitude of $\|\varDelta\mathbf{D}_k\|_F$ is below certain threshold, so that the gap between the two domains is absorbed into the learned intermediate domain dictionaries. This stopping criteria also automatically gives the number of intermediate domains to sample from the transition path. We summarize our approach in Algorithm 1.

1: Input:Dictionary $\mathbf{D}_0$ trained from the source data, target data $\mathbf{Y}_t$, sparsity level T, stopping threshold δ, parameter λ, $k = 0$.
2: Output: Dictionaries $\{\mathbf{D}_k\}_{k=1}^{K-1}$ for the intermediate domains, dictionary $\mathbf{D}_K$ for the target domain.
3: **while** stopping criteria is not reached **do**
4: Decompose the target data with the current intermediate domain dictionary $\mathbf{D}_k$, get the reconstruction residue $\mathbf{J}_k$ using equation (10),(11)
5: Get an estimate of the adjustment in dictionary atoms $\Delta\mathbf{D}_k$ and the next intermediate domain dictionary $\mathbf{D}_{k+1}$ using equation (13),(14). Normalize the atoms in $\mathbf{D}_{k+1}$ to have unit norm.
6: $k \leftarrow k + 1$
7: check the stopping criteria $\|\Delta\mathbf{D}_k\|_F \leq \delta$
8: **end while**

Algorithm 1: Algorithm to generate intermediate subspaces between source and target domains.

4.2 Recognition Under Domain Shift

To this end, we have learned a transition path which is encoded with the underlying domain shift. This provides us with rich information to obtain new representations to associate source and target data. Here, we simply apply invariant sparse codes across source, intermediate, target domain dictionaries $\{\mathbf{D}_k\}_{k=0}^{K}$. The new augmented feature representation is obtained as follows:

$$[(\mathbf{D}_0\mathbf{x})^T, (\mathbf{D}_1\mathbf{x})^T, ..., (\mathbf{D}_K\mathbf{x})^T]^T$$

where $\mathbf{x} \in \mathbb{R}^m$ is the sparse code of a source data signal decomposed with $\mathbf{D}_0$, or a target data signal decomposed with $\mathbf{D}_K$. This new representation incorporates the smooth domain transition recovered in the intermediate dictionaries into the signal space. It brings source and target data into a shared space where the data distribution shift is mitigated. Therefore, it can serve as a more robust characteristic across different domains. Given the new feature vectors, we apply random projection for dimension reduction, and then employ a SVM classifier for cross-domain recognition.

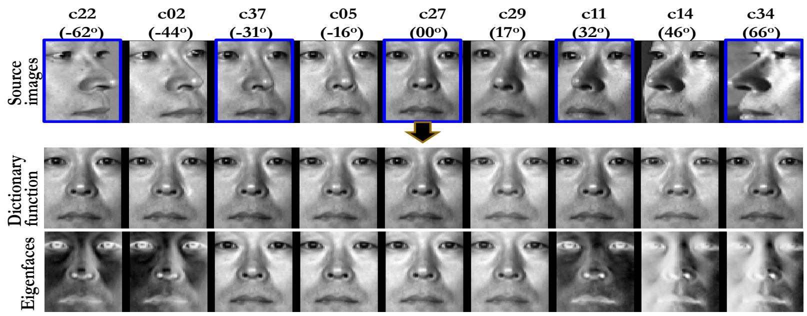


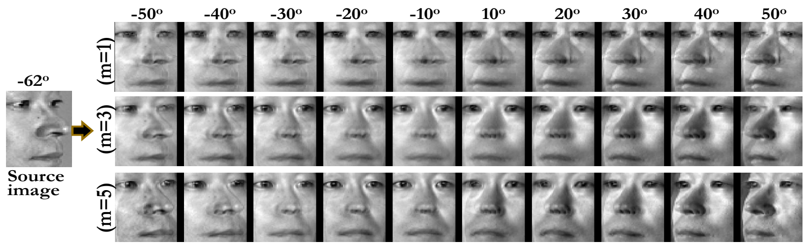
Fig. 4: Frontal face alignment. For the first row of source images, pose azimuths are shown below the camera numbers. Poses highlighted in blue are known poses to learn a linear dictionary function (m=4), and the remaining are unknown poses. The second and third rows show the aligned face to each corresponding source image using the linear dictionary function and Eigenfaces respectively.

5 Experimental Evaluation

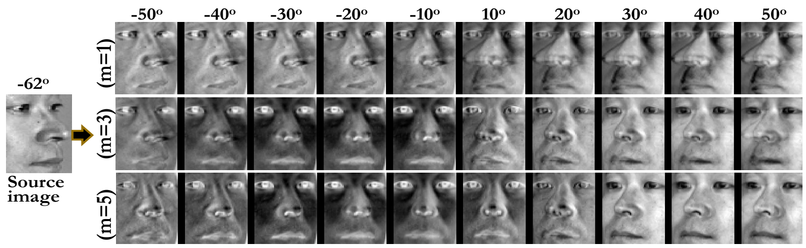
We present the results of experiments using two public face datasets: the CMU PIE dataset [29] and the Extended YaleB dataset [11]. The CMU PIE dataset consists of 68 subjects in 13 poses and 21 lighting conditions. In our experiments we use 9 poses which have approximately the same camera altitude, as shown in the first row of Figure 4. The Extended YaleB dataset consists of 38 subjects in 64 lighting conditions. All images are in 64×48 size. We will first evaluate the basic behaviors of DADL through pose alignment. Then we will demonstrate the effectiveness of both DADL and UDADL in face re-identification across domain.

5.1 DADL for pose alignment

Frontal Face Alignment: In Figure 4, we align different face poses to the frontal view. We learn for each subject in the PIE dataset a linear dictionary function $F(\boldsymbol{\theta}, \mathbf{W})$ (m=4) using 5 out of 9 poses. The training poses are highlighted in blue in the first row of Figure 4. Given a source image $\mathbf{y_s}$, we first estimate the domain parameters $\boldsymbol{\theta_s}$, i.e., the pose azimuth here, as discussed in [26]. We then obtain the sparse representation $\mathbf{x_s}$ of the source image as $\min_{\mathbf{x_s}} \|\mathbf{y_s} - F(\boldsymbol{\theta_s}, \mathbf{W})\mathbf{x_s}\|_2^2,\ s.t.\ \|\mathbf{x_s}\|_o \leq T$ (sparsity level) using any pursuit methods such as OMP [10]. We specify the frontal pose azimuth (00^o) as the parameter for the target domain $\boldsymbol{\theta_t}$, and obtain the frontal view image $\mathbf{y_t}$ as $\mathbf{y_t} = F(\boldsymbol{\theta_t}, \mathbf{W})\mathbf{x_s}$. The second row of Figure 4 shows the aligned frontal view images to the respective poses in the first row. These aligned frontal faces are close to the actual image, i.e., c27 in the first row. It is noted that images with poses c02, c05, c29 and c14 are unknown poses to the learned dictionary function.



(a) Pose synthesis using a linear dictionary function

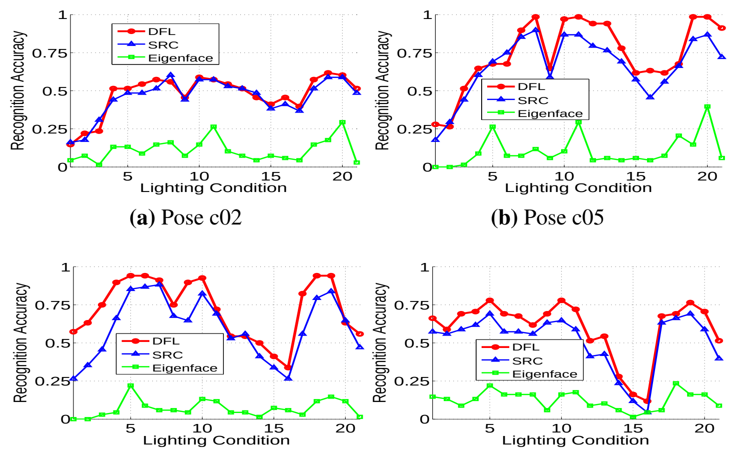


(b) Pose synthesis using Eigenfaces

Fig. 5: Pose synthesis using various degrees of dictionary polynomials. All the synthesized poses are unknown to learned dictionary functions and associated with no actual observations. m is the degree of a dictionary polynomial in (6).

For comparison purposes, we learn Eigenfaces for each of the 5 training poses and obtain adapted Eigenfaces at 4 unknown poses using the same function fitting method in our framework. We then project each source image (mean-subtracted) on the respective Eigenfaces and use frontal Eigenfaces to reconstruct the aligned image shown in the third row of Figure 4. The proposed method of jointly learning the dictionary function parameters and domain-invariant sparse codes in (8) significantly outperforms the Eigenfaces approach, which fails for large pose variations.

Pose Synthesis: In Figure 5, we synthesize new poses at any given pose azimuth. We learn for each subject in the PIE dataset a linear dictionary function $F(\boldsymbol{\theta}, \mathbf{W})$ using all 9 poses. In Figure 5a, given a source image $\mathbf{y_s}$ in a profile pose (-62^o), we first estimate the domain parameters $\boldsymbol{\theta_s}$ for the source image, and sparsely decompose it over $F(\boldsymbol{\theta_s}, \mathbf{W})$ for its sparse representation $\mathbf{x_s}$. We specify every 10^o pose azimuth in $[-50^o, 50^o]$ as parameters for the target domain $\boldsymbol{\theta_t}$, and obtain a synthesized pose image $\mathbf{y_t}$ as $\mathbf{y_t} = F(\boldsymbol{\theta_t}, \mathbf{W})\mathbf{x_s}$. It is noted that none of the target poses are associated with actual observations. As shown in Figure 5a, we obtain reasonable synthesized images at poses with no observations. We observe improved synthesis performance by increasing the value of m, i.e., the degree of a dictionary polynomial. In Figure 5b, we perform curve fitting over Eigenfaces as discussed. The proposed dictionary function learning framework exhibits better synthesis performance.



(a) Pose c02 **(b)** Pose c05

(c) Pose c29 **(d)** Pose c14

Fig. 6: Face recognition accuracy on the CMU PIE dataset. The proposed method is denoted as DFL in color red.

5.2 DADL for face re-identification

Two face recognition methods are adopted for comparisons: Eigenfaces [31] and SRC [34]. SRC is a state of the art method to use sparse representation for face recognition. We denote our method as the Dictionary Function Learning (DFL) method. For a fair comparison, we adopt exactly the same configurations for all the three methods, i.e., we use 68 subjects in 5 poses c22, c37, c27, c11 and c34 in the PIE dataset for training, and the remaining 4 poses for testing.

For the SRC method, we form a dictionary from the training data for each pose of a subject. For the proposed DFL method, we learn from the training data a dictionary function across pose for each subject. In SRC and DFL, a test image is classified using the subject label associated with the dictionary or the dictionary function respectively that gives the minimal reconstruction error. In Eigenfaces, a nearest neighbor classifier is used. In Figure 6, we present the face recognition accuracy on the PIE dataset for different testing poses under each lighting condition. The proposed DFL method outperforms both Eigenfaces and SRC methods for all testing poses.

5.3 Unsupervised DADL for face re-identification

Across pose variation: We present the results of face recognition across pose variation using the CMU-PIE dataset [29]. This experiment includes 68 subjects under 5 different poses. Each subject has 21 images at each pose, with variations in lightings. We select the frontal face images as the source domain, with a total of 1428 images. The target domain contains images at different poses, which are denoted as $c05$ and $c29$ (yawning about $\pm 22.5^o$), $c37$ and $c11$ (yawning bout $\pm 45^o$) respectively. We choose the front-illuminated source images to be the labeled data in the source domain. The task is to determine the identity of faces in the target domain with the same illumination condition. The classification results are in Table 1. We compare our method with the following methods. 1) Baseline K-SVD [1], where the target data is represented using the dictionary learned from the source domain, and the resulting sparse codes are compared using nearest neighbor classifier. 2) GFK [12] and SGF [13], which perform subspace interpolation via infinite or finite sampling on the Grassmann manifold. 3) Eigen light-field [14] method, which is specifically designed to handle face recognition across pose variations. We observe that the baseline is heavily biased under domain shift, and all the DA methods improve upon it. Our method has advantages over other two DA methods when the pose variation is large. Further, our average performance is competitive with [14], which relies on a generic training set to build pose specific models, while DA methods do not make such an assumption. We also show some of the synthesized intermediate images in Figure 7 for an illustration. As our DA approach gradually updates the dictionary learned from frontal face images using non-frontal images, these transformed representations thus convey the transition process in this scenario. These transformations could also provide additional information for certain applications, e.g. face reconstruction across different poses.

Table 1: Face recognition under pose variation on CMU-PIE dataset [29]

	c11	c29	c05	c37	average
Ours	76.5	**98.5**	**98.5**	88.2	**90.4**
GFK [12]	63.2	92.7	92.7	76.5	81.3
SGF [13]	51.5	82.4	82.4	67.7	71.0
Eigen light-field [14]	**78.0**	91.0	93.0	**89.0**	87.8
K-SVD [1]	48.5	76.5	80.9	57.4	65.8

Across blur and illumination variations: Next we performed a face recognition experiment across combined blur and illumination variations. All frontal images of the first 34 subjects under 21 lighting conditions from the CMU-PIE dataset [29] are included in this experiment. We randomly select images under 11 different illumination conditions to form the source domain. The remaining images with the other 10 illumination conditions are convolved with a blur kernel to form the target domain. Experiments are performed with the Gaussian kernels with standard deviations of 3 and 4, and motion blurs with lengths of 9 (angle $\theta = 135^o$) and 11 (angle $\theta = 45^o$), respectively. We compare our results with those of K-SVD [1], GFK [12], and SGF [13]. Besides, we also compare with the Local Phase Quantization [2] method, which is a blur insensitive descriptor, and the method based in [3], which estimates an albedo map (Albedo) as an illumination robust signature for

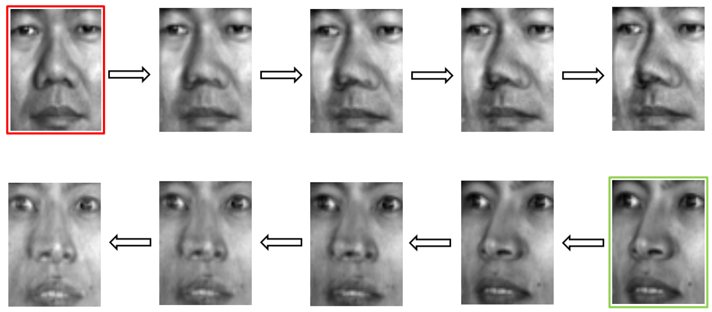


Fig. 7: Synthesized intermediate representations between frontal face images and face images at pose $c11$. The first row shows the transformed images from a source image (in red box) to the target domain. The second row shows the transformed images from a target image (in green box) to the source domain.

matching. We report the results in Table 2. Our method is competitive with [12], and outperforms all other algorithms by a large margin. Since the domain shift in this experiment consists of both illumination and blur variation, traditional methods which are only illumination insensitive or robust to blur are not able to fully handle both variations. DA methods are useful in this scenario as they do not rely on the knowledge of physical domain shift. We also show transformed intermediate representations along the transition path of our approach in Figure 8, which clearly captures the transition from clear to blur images and vice versa. Particularly, we believe that the transformation from blur to clear conditions is useful for blind deconvolution, which is a highly under-constrained and costly problem [17].

Table 2: Face recognition across illumination and blur variations on CMU-PIE dataset [29]

	$\sigma = 3$	$\sigma = 4$	$L = 9$	$L = 11$
Ours	**80.29**	**77.94**	**85.88**	**81.18**
GFK [12]	78.53	77.65	82.35	77.65
SGF [13]	63.82	52.06	70.29	57.06
LPQ [2]	66.47	32.94	73.82	62.06
Albedo [3]	50.88	36.76	60.88	45.88
K-SVD [1]	40.29	25.59	42.35	30.59

6 Conclusions

In this chapter, we presented two different methods for the face re-identification problem using the domain adaptive dictionary learning approach. We first presented

Fig. 8: Synthesized intermediate representations from the experiment on face recognition across illumination and blur variations (motion blur with length of 9). The first row demonstrates the transformed images from a source image (in red box) to the target domain. The second row demonstrates the transformed images from a target image (in green box) to the source domain.

a general dictionary function learning framework to transform a dictionary learned from one domain to the other. Domain dictionaries are modeled by a parametric function. The dictionary function parameters and domain-invariant sparse codes are then jointly learned by solving an optimization problem with a sparsity constraint. We then discussed a fully unsupervised domain adaptive dictionary learning method with no prior knowledge of the underlying domain shift. This unsupervised DA method learns a set of intermediate domain dictionaries between the source and target domains, and render intermediate domain representations to form a shared feature space for re-identification of faces. Extensive experiments on real datasets demonstrate the effectiveness of these methods on applications such as face pose alignment and face re-identification across domains.

Acknowledgements

The work reported here is partially supported by a MURI Grant N00014-08-1-0638 from the Office of Naval Research.

References

1. Aharon, M., Elad, M., Bruckstein, A.: K-svd : An algorithm for designing of overcomplete dictionaries for sparse representation. IEEE Transactions on Signal Processing **54**(11), 4311–4322 (2006)
2. Ahonen, T., Rahtu, E., Ojansivu, V., Heikkilä, J.: Recognition of blurred faces using local phase quantization. In: International Conference on Pattern Recognition (2008)
3. Biswas, S., Aggarwal, G., Chellappa, R.: Robust estimation of albedo for illumination-invariant matching and shape recovery. IEEE Transactions on Pattern Analysis and Machine

Intelligence **31**, 884–899 (2009)
4. Blitzer, J., McDonald, R., Pereira, F.: Domain adaptation with structural correspondence learning. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (2006)
5. Chellappa, R., Ni, J., Patel, V.M.: Remote identification of faces: Problems, prospects, and progress. Pattern Recogn. Lett. **33**, 1849–1859 (2012)
6. Chen, S., Donoho, D., Saunders, M.: Atomic decomposition by basis pursuit. SIAM J. Sci. Comp. **20**, 33–61 (1998)
7. Daume III, H.: Frustratingly easy domain adaptation. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (2007)
8. Duan, L., Xu, D., Tsang, I.W.H., Luo, J.: Visual event recognition in videos by learning from web data. IEEE Transactions on Pattern Analysis and Machine Intelligence **99** (2011)
9. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image Processing **15**(12), 3736–3745 (2006)
10. Engan, K., Aase, S.O., Hakon Husoy, J.: Method of optimal directions for frame design. In: International Conference on Acoustics, Speech, and Signal Processing (1999)
11. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**, 643–660 (2001)
12. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: Proc. IEEE Computer Society Conf. on Computer Vision and Patt. Recn. IEEE (2012)
13. Gopalan, R., Li, R., Chellappa, R.: Domain adaptation for object recognition: An unsupervised approach. In: International Conference on Computer Vision (2011)
14. Gross, R., Matthews, I., Baker, S.: Appearance-based face recognition and light-fields. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**, 449–465 (2004)
15. Jhuo, I.H., Liu, D., Lee, D.T., Chang, S.F.: Robust visual domain adaptation with low-rank reconstruction. In: Proc. IEEE Computer Society Conf. on Computer Vision and Patt. Recn. IEEE (2012)
16. Kulis, B., Saenko, K., Darrell, T.: What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In: Proc. IEEE Computer Society Conf. on Computer Vision and Patt. Recn. IEEE (2011)
17. Levin, A., Weiss, Y., Durand, F., Freeman, W.T.: Understanding and evaluating blind deconvolution algorithms. In: Proc. IEEE Computer Society Conf. on Computer Vision and Patt. Recn. IEEE (2009)
18. Machado, L., Leite, F.S.: Fitting smooth paths on riemannian manifolds. Int. J. Appl. Math. Stat. **4**, 25–53 (2006)
19. Mairal, J., Elad, M., Sapiro, G.: Sparse representation for color image restoration. IEEE Transactions on Image Processing **17**(1), 53–69 (2008)
20. Ni, J., Qiu, Q., Chellappa, R.: Subspace interpolation via dictionary learning for unsupervised domain adaptation. In: Proc. IEEE Computer Society Conf. on Computer Vision and Patt. Recn. IEEE (2013)
21. Olshausen, B.A., Field, D.J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature **381**(6583), 607–609 (1996)
22. Pan, S.J., Kwok, J.T., Yang, Q.: Transfer learning via dimensionality reduction. In: Proceedings of the 23rd national conference on Artificial intelligence (2008)
23. Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q.: Domain adaptation via transfer component analysis. In: International Joint Conferences on Artificial Intelligence (2009)
24. Pati, Y.C., Rezaiifar, R., Krishnaprasad, P.S.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. Proc. 27th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA pp. 40–44 (1993)
25. Qiu, Q., Jiang, Z., Chellappa, R.: Sparse dictionary-based representation and recognition of action attributes. In: International Conference on Computer Vision, pp. 707–714. IEEE (2011)
26. Qiu, Q., Patel, V., Turaga, P., Chellappa, R.: Domain adaptive dictionary learning. In: Proc. European Conference on Computer Vision (2012)

27. Rubinstein, R., Bruckstein, A., Elad, M.: Dictionaries for sparse representation modeling. Proceedings of the IEEE **98**(6), 1045 –1057 (2010)
28. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting visual category models to new domains. In: Proc. European Conference on Computer Vision (2010)
29. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (PIE) database. IEEE Transactions on Pattern Analysis and Machine Intelligence **25**(12), 1615–1618 (2003)
30. Tropp, J.: Greed is good: algorithmic results for sparse approximation. IEEE Trans. on Information Theory **50**, 2231 – 2242 (2004)
31. Turk, M., Pentland, A.: Face recognition using eigenfaces. In: Proc. IEEE Computer Society Conf. on Computer Vision and Patt. Recn. (1991)
32. Wang, C., Mahadevan, S.: Manifold alignment without correspondence. In: International Joint Conferences on Artificial Intelligence, pp. 1273–1278 (2009)
33. Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T., Yan, S.: Sparse representation for computer vision and pattern recognition. Proceedings of the IEEE **98**(6), 1031–1044 (2010)
34. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**, 210–227 (2009)
35. Yang, J., Yan, R., Hauptmann, A.G.: Cross-domain video concept detection using adaptive svms. In: ACM Multimedia, pp. 188–197. ACM (2007)
36. Zhang, Q., Li, B.: Discriminative k-svd for dictionary learning in face recognition. In: Proc. IEEE Computer Society Conf. on Computer Vision and Patt. Recn. IEEE (2010)
37. Zheng, W.S., Gong, S., Xiang, T.: Reidentification by relative distance comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**, 653–668 (2013)

Index

Blur variation, 13

Dictionary atom, 4
Dictionary learning, 2
Domain adaptation, 2
Domain adaptive dictionary learning, 2, 5
Domain dictionary function, 6
Domain dictionary function learning, 6
Domain shift, 9

Face re-identification, 2, 12, 13

Illumination variation, 13
Intermediate domain dictionary, 8

K-SVD, 4

Pose alignment, 10
Pose variation, 13

Source domain, 2
Sparse code, 4
Sparse Representation, 3
Sparsity, 4

Target domain, 2

Unsupervised domain adaptive dictionary learning, 3, 7

Vector transpose, 5