Domain adaptation (DA) tackles the problem where data from the training set (source domain) and test set (target domain) have different underlying distributions. For instance, training and testing images may be acquired under different environments, viewpoints and illumination conditions. In this paper, we focus on the more challenging unsupervised DA problem where the samples in the target domain are unlabeled. It is noticed that dictionary learning has gained a lot of popularity due to the fact that images of interest could be reconstructed sparsely in an appropriately learned dictionary [1].

Specifically, we propose a novel domain-adaptive dictionary learning approach to generate a set of intermediate domains which bridge the gap between source and target domains. Our approach defines two types of dictionaries: a common dictionary and a domain-specific dictionary. The overall learning process illustrated in Figure 1 consists of three steps: (1) At the beginning, we first learn the common dictionary $D^f$, domain-specific dictionaries $D^0$ and $D^f$ for source and target domains. (2) At the $k$-th step, we enforce the recovered feature representations of target data in all available domains to have the same sparse codes, while adapting the most recently obtained dictionary $D^k$ to better represent the target domain. Then we multiply dictionaries in the $k$-th domain with the corresponding sparse codes to recover feature representations of target data $X^t_1$ in this domain. (3) We update $D^k$ to find the next domain-specific dictionary $D^{k+1}$ by further minimizing the reconstruction error in representing the target data. Then we alternate between the steps of sparse coding and dictionary updating until the stopping criteria is satisfied.

**Notations:** Let $X_o \in \mathbb{R}^{d \times N_s}$, $X_t \in \mathbb{R}^{d \times N_t}$ be the feature representations of source and target data respectively, where $d$ is the feature dimension, $N_s$ and $N_t$ are the number of samples in the two domains. The feature representations of recovered source and target data in the $k$-th intermediate domain are denoted as $X^D_k \in \mathbb{R}^{d \times N_s}$ and $X^{tD}_k \in \mathbb{R}^{d \times N_t}$ respectively. The common dictionary is denoted as $D^f$, whereas source-specific and target-specific dictionaries are denoted as $D^0$, $D^f$ respectively. Similarly, we use $D^k$, $k = 1...N_t$ to denote the domain-specific dictionary for the $k$-th domain, where $N$ is the number of intermediate domains. We set all the dictionaries to be of the same size $\mathbb{R}^{d \times N}$. At the beginning, we learn the common dictionary $D^f$ by minimizing the reconstruction error of both source and target data as follows:

$$
\min_{D^f, D^0} \|X_o - D^f Z^0\|_F^2 + \|X_t - D^f Z^0\|_F^2 \quad \text{s.t.} \quad \|Z^0_o\|_0 \leq T, \|Z^0_t\|_0 \leq T
$$

where $D^0 = [d_0^\top,...,d_{N_s}^\top] \in \mathbb{R}^{d \times N_s}$ and $Z^0 = [z_1^\top,...,z_{N_t}^\top] \in \mathbb{R}^{d \times N_t}$ are sparse representations of $X_o$ and $X_t$ respectively, $T$ is the sparsity level.

Then we learn domain-specific dictionaries $D^0$ and $D^f$ by further reducing the reconstruction error of the source and target data as follows:

$$
\min_{D^0, D^f} \|X_o - D^f Z^0 - D^0 T^0\|_F^2 + \lambda \|D^f D^T\|_F^2 \quad \text{s.t.} \quad \|Z^0_o\|_0 + \|\alpha^0\|_0 \leq T
$$

and

$$
\min_{D^0, D^f} \|X_t - D^f Z^0 - D^0 T^f\|_F^2 + \lambda \|D^f D^T\|_F^2 \quad \text{s.t.} \quad \|Z^0_t\|_0 + \|\alpha^0\|_0 \leq T
$$

where $\Gamma^0 = [\alpha^0_1,...,\alpha^0_{N_s}] \in \mathbb{R}^{d \times N_s}$ and $\Gamma^f = [\alpha^f_1,...,\alpha^f_{N_t}] \in \mathbb{R}^{d \times N_t}$ are sparse representations of $X_o$ and $X_t$ with respect to $D^0$ and $D^f$, and $\lambda$ is the regularization parameter.

Next, we alternate between domain-adaptive sparse coding and dictionary updating to learn a set of domain-specific dictionaries for generating intermediate domains. We regularize the sparse representation of $X_o$, $X_t$ and $X^t_1$ to be the same as follows:

$$
Z^0, T^0 = \arg\min_{X_o, X_t, X^t_1} \left\{ \|X_o - D^f Z - D^0 T\|_F^2 + \sum_{i=1}^{N_t} \|X_t - D^f Z - D^0 T^f\|_F^2 \right. \\
+ \left. \|X_o - D^f Z - D^0 T^f\|_F^2 \quad \text{s.t.} \quad \|Z_o\|_0 + \|\alpha_o\|_0 \leq T, \|Z_t\|_0 + \|\alpha_t\|_0 \leq T \right\}
$$

**Source**

\[ X^0 \rightarrow \cdots \rightarrow X^k \rightarrow X^{k+1} \rightarrow \cdots \rightarrow X
\]

**Domain k**

\[ D^k Z^0 + \Delta D^k_0 \rightarrow D^k Z^1 + \Delta D^k_1 \rightarrow \cdots \rightarrow D^k Z^t + \Delta D^k_t \rightarrow \cdots \rightarrow D^k Z^N_t \]

**Domain k+1**

\[ \Delta D^k_0 \rightarrow \Delta D^k_1 \rightarrow \cdots \rightarrow \Delta D^k_t \rightarrow \cdots \rightarrow \Delta D^k_0 \]

**Target**

\[ D^f Z^0 \rightarrow D^f Z^1 \rightarrow \cdots \rightarrow D^f Z^N_t \]

Figure 1: The overall learning process in domain-adaptive dictionary learning framework.

Figure 2: The first image is the original target face image, the second image is the recovered face image corresponding to the common dictionary. The remaining six images are the recovered face images corresponding to domain-specific dictionaries.

Then after sparse coding at the $k$-th step, we update $D^k$ to find the next domain-specific dictionary $D^{k+1}$ by further reducing the reconstruction error of target data in the $k$-th domain as in [2],

$$
\min_{D^k} \|\tilde{Z}^k - D^k \tilde{T}^k\|_F^2 + \eta \|\Delta D^k\|_2^2
$$

where $\tilde{Z}^k = X_t - D^f Z^k - D^k T^k$ is the target reconstruction residue in the $k$-th domain. The next domain-specific dictionary $D^{k+1}$ is obtained as:

$$
D^{k+1} = D^k + \Delta D^k
$$

We use the sparse codes obtained in the last iterative step to derive new feature representations for the source and target data. The new augmented feature representation of source and target data are $\tilde{X}_k = [X^0_o, ..., X^k_o]$, and $X_t = [X^0_t, ..., X^k_t]$ respectively, where $\tilde{X}_k = [D^k Z^0, D^k T^0, D^k Z^k, D^k T^k]$ and $X_t = [D^k Z^0, D^k T^0, D^k Z^k, D^k T^k]$ are the sparse codes obtained in the last iterative step where $k = N$. The final stage of recognition across all the domains is performed using an SVM classifier trained on new feature vectors after dimension reduction via the Principal Component Analysis (PCA).

**Visualization:** We demonstrate the benefits of the separation of the common dictionary from domain-specific dictionaries. We select frontal face images as the source domain and face images at 45° yaw as the target domain. We visualize the recovered face images along the intermediate domains in Figure 2. Each recovered face image has two components corresponding to the learned dictionaries respectively. It can be seen that the components corresponding to domain-specific dictionaries in intermediate domains gradually adapt from the frontal face to non-frontal faces. This demonstrates that the domain-specific dictionaries have the ability to encode the domain shift caused by different yaw angles.

We also extensively evaluate our approach using the DA benchmark dataset [3] and experimental results clearly demonstrate the effectiveness of our approach.

