



Unsupervised Domain Adaptation Dictionary Learning for Visual Recognition

Zhun Zhong^(✉), Zongmin Li, Runlin Li, and Xiaoxia Sun

College of Computer and Communication Engineering,
China University of Petroleum, Qingdao, China
zhunzhong007@gmail.com

Abstract. Over the last years, dictionary learning method has been extensively applied to deal with various computer vision recognition applications, and produced state-of-the-art results. However, when the data instances of a target domain have a different distribution than that of a source domain, the dictionary learning method may fail to perform well. In this paper, we address the cross-domain visual recognition problem and propose a simple but effective unsupervised domain adaptation approach, where labeled data are only from source domain. In order to bring the original data in source and target domain into the same distribution, the proposed method forcing nearest coupled data between source and target domain to have identical sparse representations while jointly learning dictionaries for each domain, where the learned dictionaries can reconstruct original data in source and target domain respectively. So that sparse representations of original data can be used to perform visual recognition tasks. We demonstrate the effectiveness of our approach on standard datasets. Our method performs on par or better than competitive state-of-the-art methods.

1 Introduction

In the past decade, machine learning has been widely used for various computer vision applications, such as multimedia retrieval [1–3], image classification [4–9], object detection [10–13], person re-identification [14–18], etc. Traditional machine learning methods often learn a model from the training data, and then apply it to the testing data. The fundamental assumption here is that the training data and testing data have the same distribution. However, in real-world applications, it cannot always guarantee that training data share the same distribution with testing data. Therefore, it may produce very poor results when the testing data and training data have the different distributions since the training model is no longer optimal on testing data. For example, applies image classification classifier trained on amazon dataset to phone photos in real life. Face recognition model trained on frontal and well-illumination images to recognize non-frontal poses and less-illumination images. This often viewed as visual

domain adaptation problem which has been increasing interest in understanding and overcoming.

Domain adaptation aims at learning an adaptive classifier by utilizing the information between source domain with a plenty of labeled data and target domain which is collected from a different distribution. Generally, we can divide domain adaptation into two settings depending on the availability of labels in the target domain data: semi-supervised domain adaptation, and unsupervised domain adaptation. In scenario of semi-supervised domain adaptation, labeled data is available in both source domain (with a plenty of labeled data) and target domain (with a few labeled data), while in scenario of unsupervised domain adaptation labeled data are only available from source domain. In this paper, we mainly focus on unsupervised domain adaptation which is a more challenging task, and more in line with the real-world applications.

Many recent works [19–21] focus on subspace based method to tackle visual domain adaptation problems. In [21], Li et al. determined a feature subspace via canonical correlation analysis (CCA) [22] for recognizing faces with different poses. In [19], Gopalan et al. using geodesic flows to generate intermediate subspaces along the geodesic path between source domain subspace and target domain subspace on the Grassmann manifold. In [20], Gong et al. proposed Geodesic Flow Kernel (GFK), which computes a symmetric kernel between source and target points based on geodesic flow along a latent manifold.

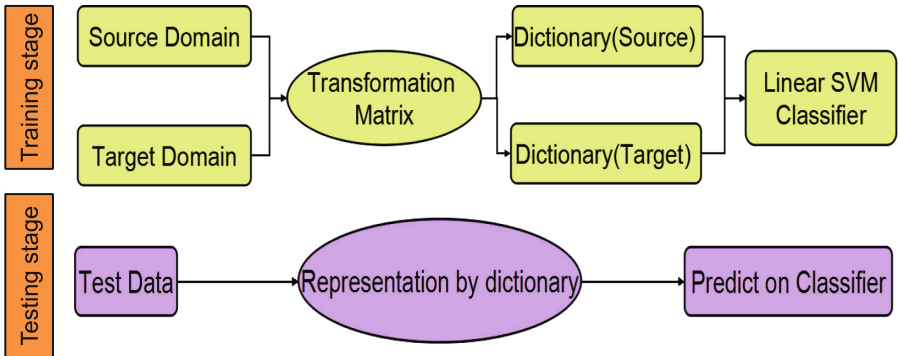


Fig. 1. The overall schema of the proposed framework.

In last few years, the study of dictionary learning based sparse representation has received extensive attention. It has been successfully used for a variety of computer vision applications. For example, classification [23], recognition [24] and denoising [25]. Using an over-complete dictionary, signal or image can be approximated by the combination of only a few number of atoms, that are chosen from the learned dictionary. One of the early dictionary learning algorithms was proposed by Olshausen and Field [26], where a maximum likelihood (ML) learning method was used to sparsely encode images upon a redundant dictionary.

Based on the same ML objective function as in [26], Engan et al. [27] developed a more efficient algorithm, called the method of optimal directions (MOD), in which a closed-form solution for the dictionary update has been proposed. More recently, in [28], Aharon, Elad and Bruckstein proposed the K-SVD algorithm by generalizing k-means clustering and efficiently learns an over-complete dictionary from a set of training signals. This method has been implemented in a variety of image processing problems.

The most existing dictionary based methods assuming that training data and testing data come from the same distribution. However, the learned dictionary may not be optimal if the testing data has different distribution from the data used for training. Learning dictionaries under different domain is a challenging task, and gradually become a hot research over the last few years. In [29], Jia et al. considered a special case where corresponding samples from each domain were available, and learn a dictionary for each domain. Qiu et al. [30] presented a general joint optimization function that transforms a dictionary learned from one domain to the other, and applied such a framework to applications such as pose alignment, pose illumination estimation, and face recognition. Zheng et al. [31] proposed a method achieved promising results on the cross-view action recognition problem with pairwise dictionaries constructed using correspondences between the target view and the source view. In [32], Shekhar et al. learn a latent dictionary which can succinctly represent both the domains in a common projected low-dimensional space. Ni et al. [33] learn a set of subspaces through dictionary learning to mitigate the divergence of source and target domains. Huang and Wang [34] proposed a joint model which learns a pair of dictionaries with a feature space for describing and associating cross-domain data. In [35, 36], Zhu and Shao proposed a weakly-supervised framework learns a pairwise dictionaries and a classifier while considering the capacity of the dictionaries in terms of reconstructability, discriminability and domain adaptability.

In this paper, we present an unsupervised domain adaptation approach through dictionary learning. Different from above dictionary learning based domain adaptation methods, our method directly learning adaptive dictionaries in low-level feature space and with no need for labels either in source domain or target domain during dictionary learning process. Our method is inspired by [35, 36], which forcing the similar samples in the same class to have identical representations in the sparse space. However, our method is unsupervised, we assume that the nearest coupled low-level features in the original space should maintain their relationship in the sparse space (i.e. these coupled features have the same sparse representation). According to this main idea, we learn a transformation matrix, which selected the nearest data in source domain to each target data. Then the dictionaries for each domain are jointly learned by these selected source data and target data. The data from each domain can be encoded by their dictionaries and then represented by sparse features. Thus, SVM classifier can be trained using these sparse features, and predicting test data on the learned classifier. The learning framework is performed by a classic and efficient dictionary learning method, K-SVD [28]. We demonstrate the effectiveness of our

approach on standard cross-domain datasets, and we get state-of-the-art results. An overall schema of the proposed framework is shown in Fig. 1.

1.1 Organization of the Paper

The structure of the rest of the paper is as follows: In Sect. 2, we present our unsupervised domain adaptation dictionary learning algorithm and introduce the classification scheme for the learned dictionary. Experimental results on object recognition are presented in Sect. 3. Finally, the conclusion of this work is given in Sect. 4.

2 Proposed Method

2.1 Problem Notation

Let $I_s = \{I_{s,i}\}_{i=1}^{N_s}$, and $I_t = \{I_{t,j}\}_{j=1}^{N_t}$ be the data instances from the source and target domain respectively, where N_s and N_t denote the number of samples. Each sample from I_s and I_t has a set of d -dimensional local features, thus each sample can be represented by $I_{s,i} = \{I_{s,i}^1, I_{s,i}^2, \dots, I_{s,i}^{M_i}\}$ and $I_{t,j} = \{I_{t,j}^1, I_{t,j}^2, \dots, I_{t,j}^{M_j}\}$ in source and target domain respectively, where M_i and M_j denote the number of local features. Then, the set of local features of source and target domain can be denoted as $Y_s \in \mathbb{R}^{d \times L_s}$, and $Y_t \in \mathbb{R}^{d \times L_t}$ respectively, where L_s and L_t denote the number of local features in the source and target domain.

2.2 Dictionary Learning

Here, we give a brief review of classical dictionary learning approach. Given a set of d -dimensional input signals, $Y \in \mathbb{R}^{d \times L}$, where L is denoted as the number of input signals. Then, learning a K -atoms dictionary of the signals Y , $D \in \mathbb{R}^{d \times K}$, can be obtained by solving the following optimization problem:

$$\begin{aligned} \{D, X\} = \operatorname{argmin}_{D, X} \|Y - DX\|_F^2 \\ \text{s.t. } \forall_i, \|x_i\|_0 \leq T_0 \end{aligned} \quad (1)$$

where $D = [d_1, d_2, \dots, d_K] \in \mathbb{R}^{d \times K}$ denotes the dictionary, $X = [x_1, x_2, \dots, x_L] \in \mathbb{R}^{K \times L}$ denotes the sparse coefficients of Y decomposed with D , and T_0 is the sparsity level that constraint the number of nonzero entries in x_i .

The performance of sparse representation strictly lie on dictionary learning method. The K-SVD algorithm [28] is a highly effective dictionary learning method that focuses on minimizing the reconstruction error. In this paper, we will solve our formulation of unsupervised domain adaptation dictionary learning based on the K-SVD algorithm.

2.3 Unsupervised Domain Adaptation Dictionary Learning

Now, consider a more general scenario, where we have data from two domains, source domain $Y_s \in \mathbb{R}^{d^*L_s}$, and target domain $Y_t \in \mathbb{R}^{d^*L_t}$. We wish to jointly learning corresponding dictionaries for each domain. Formally, we desire to minimize the following cost function:

$$\begin{aligned} & \{D_s, D_t, X_s, X_t\} \\ & = \operatorname{argmin}_{D_s, D_t, X_s, X_t} \|Y_s - D_s X_s\|_F^2 \\ & + \|Y_t - D_t X_t\|_F^2 \quad \text{s.t.} \quad \forall_i, [\|x_i^s\|_0, \|x_i^t\|_0] \leq T_0 \end{aligned} \quad (2)$$

In addition, in order to maintain the relationship in original feature space, we assume that the nearest coupled low-level features in the original space should also be the nearest couple in the sparse space. Now the new cost function is given by:

$$\begin{aligned} & \{D_s, D_t, X_s, X_t\} \\ & = \operatorname{argmin}_{D_s, D_t, X_s, X_t} \|Y_s - D_s X_s\|_F^2 \\ & + \|Y_t - D_t X_t\|_F^2 + C([X_s X_t]) \\ & \text{s.t.} \quad \forall_i, [\|x_i^s\|_0, \|x_i^t\|_0] \leq T_0 \end{aligned} \quad (3)$$

where $D_s = [d_1^s, d_2^s, \dots, d_K^s] \in \mathbb{R}^{d^*K}$ is the learned source domain dictionary, $X_s = [x_1^s, x_2^s, \dots, x_{L_s}^s] \in \mathbb{R}^{K^*L_s}$ is the sparse coefficients of source domain, $D_t = [d_1^t, d_2^t, \dots, d_K^t] \in \mathbb{R}^{d^*K}$ is the learned target domain dictionary, and $X_t = [x_1^t, x_2^t, \dots, x_{L_t}^t] \in \mathbb{R}^{K^*L_t}$ is the sparse coefficients of target domain. The function $C(\cdot)$ is defined as the distance in the new sparse space of original nearest couples, a small $C(\cdot)$ indicates the data maintain more relationship in new sparse space. This idea is inspired by [35, 36], in their method, this function is designed to measure the distances of similar cross-domain instances of the same class. However, our method is exactly unsupervised and directly perform on low-level feature. Thus, the function $C([X_s X_t])$ is defined as:

$$C([X_s X_t]) = \|X_t - X_s P\|_F^2 \quad (4)$$

where $P \in \mathbb{R}^{L_s^*L_t}$ is the transformation matrix which records the nearest couples between the original data in source and target domain, P can be represented by:

$$P = \begin{pmatrix} \Phi(y_1^s, y_1^t) & \dots & \dots & \Phi(y_1^s, y_{L_t}^t) \\ \vdots & & \ddots & \vdots \\ \vdots & & \ddots & \vdots \\ \Phi(y_{L_s}^s, y_1^t) & \dots & \dots & \Phi(y_{L_s}^s, y_{L_t}^t) \end{pmatrix} \quad (5)$$

where $\Phi(y_i^s, y_j^t)$ is the Gaussian distance between data in original feature space:

$$\Phi(y_i^s, y_j^t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y_i^s - y_j^t}{2}} \quad (6)$$

Then, P can be computed by selecting the maximum entry in each column and set to 1 while the other entries are set to 0:

$$P = (i, j) = \begin{cases} 1 & \text{if } P(i, j) = \max(P(:, j)) \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Thus, Eq. (3) can be written as:

$$\begin{aligned} & \{D_s, D_t, X_s, X_t\} \\ & = \operatorname{argmin}_{D_s, D_t, X_s, X_t} \|Y_s - D_s X_s\|_F^2 \\ & \quad + \|Y_t - D_t X_t\|_F^2 + \|X_t - X_s P\|_F^2 \\ & \text{s.t. } \forall_i, [\|x_i^s\|_0 \|x_i^t\|_0] \leq T_0 \end{aligned} \quad (8)$$

Assuming P leads to a perfect mapping across the sparse codes X_t and X_s , and each nearest couple has an identical representation after encoding, then $\|X_t - X_s P\|_F^2 = 0$. Thus $X_t = X_s P$, we can rewritten Eq. (8) as:

$$\begin{aligned} & \{D_s, D_t, X_s, X_t\} \\ & = \operatorname{argmin}_{D_s, D_t, X_s, X_t} \|(Y_s - D_s X_s)P\|_F^2 + \|Y_t - D_t X_t\|_F^2 \\ & = \operatorname{argmin}_{D_s, D_t, X_s, X_t} \|Y_s P - D_s X_s P\|_F^2 + \|Y_t - D_t X_t\|_F^2 \\ & = \operatorname{argmin}_{D_s, D_t, X_s, X_t} \|Y_s P - D_s X_t\|_F^2 + \|Y_t - D_t X_t\|_F^2 \\ & \text{s.t. } \forall_i, \|x_i^t\|_0 \leq T_0 \end{aligned} \quad (9)$$

2.4 Optimization

We can written Eq. (9) as:

$$\begin{aligned} \{\tilde{D}, \tilde{X}\} & = \operatorname{argmin}_{\tilde{D}, \tilde{X}} \|\tilde{Y} - \tilde{D}\tilde{X}\|_F^2 \\ & \text{s.t. } \forall_i, \|\tilde{x}_i\|_0 \leq T_0 \end{aligned} \quad (10)$$

where $\tilde{Y} = \begin{pmatrix} Y_s P \\ Y_t \end{pmatrix}$, $\tilde{D} = \begin{pmatrix} D_s \\ D_t \end{pmatrix}$, and $\tilde{X} = X_t$. Thus, such optimization problem can be solved using the K-SVD algorithm [28].

2.5 Object Recognition

Given the learned D_s and D_t , we obtain sparse representations of the training data in source domain and testing data in target domain respectively. For each image, we obtain a set of sparse representation $X_i = [x_{i,1}, x_{i,2}, \dots, x_{i,M_i}] \in \mathbb{R}^{K \times M_i}$, where $X_{i,j}$ is the sparse representation of j^{th} feature in image i , K denotes the dictionary size, and M_i is the number of local feature in image i . Then each image represented by a K -vector global representation through max pooling the sparse codes of local features, and then we use linear SVM classifier for cross-domain recognition.



Fig. 2. Example images from the LAPTOP category on four datasets.

3 Experiments

In this section, we evaluate our domain adaptation approach on 2D object recognition across different datasets.

Experimental Setup: Following the experiment setting in [20], we evaluate our domain adaptation approach on four datasets: Amazon (images downloaded from online merchants), Webcam (low resolution images by a web camera), Dslr (high-resolution images by a SLR camera), and Caltech-256 [37]. We regard each dataset as a domain. Figure 2 shows sample images from these datasets, and clearly highlights the differences between them. We extract 10 classes common to all four datasets: BACKPACK, TOURING-BIKE, CALCULATOR, HEADPHONES, COMPUTER-KEYBOARD, LAPTOP-101, COMPUTER-MONITOR, COMPUTER-MOUSE, COFFEEMUG, AND VIDEO-PROJECTOR. There are 2533 images in total. Each class has 8 to 151 images in a dataset. We use a SURF detector [38] to extract local features over all images. For each pair of source and target domains, we use 20 training samples per class for Amazon/Caltech, and 8 samples per class for DSLR/Webcam when used as source. To draw complete comparison with existing domain adaptation methods, we also carried out experiments on the semi-supervised setting where we additionally sampled 3 labeled images per class from the target domain. We ran 20 different trials corresponding to different selections of labeled data from the source and target domains and testing all unlabeled data in target domain. Our baseline is BOW, where all the images were represented by 800-bin histograms over the codebooks trained from a subset of Amazon images. Our

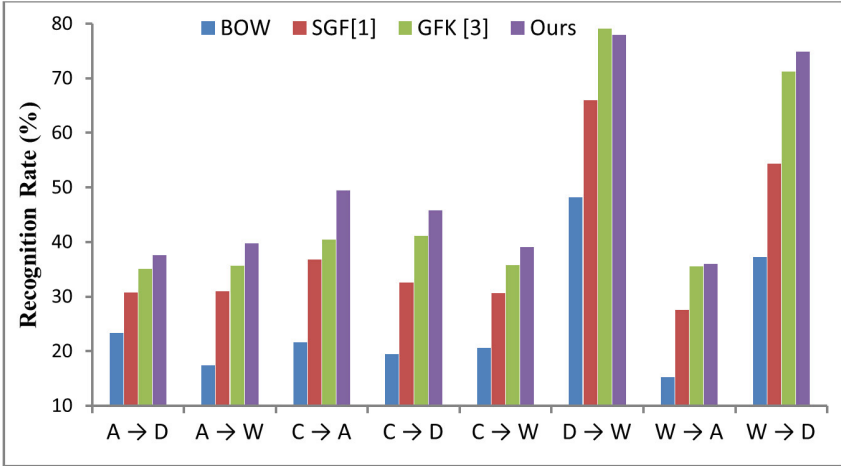


Fig. 3. Cross dataset object recognition accuracies on target domains with unsupervised adaptation over the four datasets (A: Amazon, C: Caltech, D: Dslr, W: Webcam).

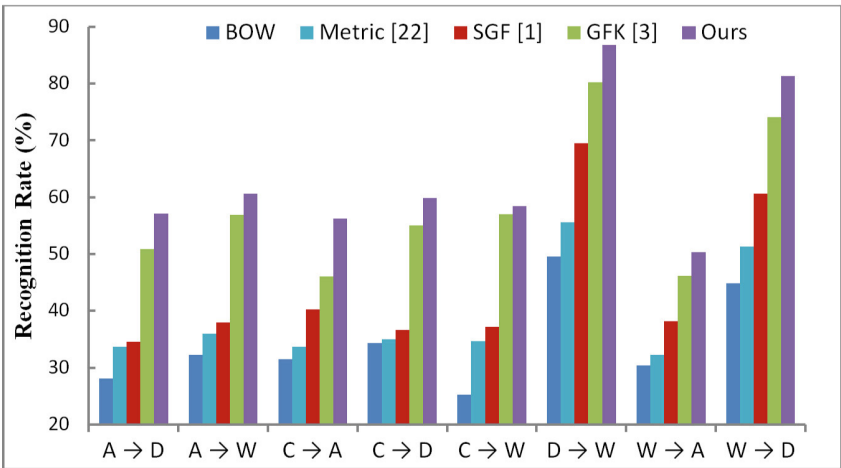


Fig. 4. Cross dataset object recognition accuracies on target domains with semi-supervised adaptation over the four datasets (A: Amazon, C: Caltech, D: Dslr, W: Webcam).

method is also compared with Metric [39], SGF [19] and GFK [20]. Note that, Metric [39] is limited to the semi-supervised setting.

Parameter Settings: For our method, we set dictionary size $K = 512$, and sparse level $T_0 = 5$ for each domain.

Results: The average recognition rate is reported in Figs. 3 and 4 for unsupervised and supervised settings respectively. It is seen that the baseline BOW has

the lowest recognition rate, all domain adaptation methods improve accuracy over it. Furthermore, GFK [20] based method clearly outperforms Metric [39] and SGF [19]. Overall, our method consistently demonstrates better performance over all methods except for one pair of source and target combination a little less than GFK [20] in the unsupervised setting.

4 Conclusions

In this paper, we presented a fully unsupervised domain adaptation dictionary learning method to jointly learning domain dictionaries by capturing the relationship between the source and target domain in the original data space. We evaluated our method on publicly available datasets and obtain improved performance upon the state of the art.

References

1. Zheng, L., Wang, S., Liu, Z., Tian, Q.: Fast image retrieval: query pruning and early termination. *IEEE Trans. Multimed.* **17**(5), 648–659 (2015)
2. Zheng, L., Wang, S., Tian, Q.: Coupled binary embedding for large-scale image retrieval. *IEEE Trans. Image Process.* **23**(8), 3368–3380 (2014)
3. Kuang, Z., Li, Z., Jiang, X., Liu, Y., Li, H.: Retrieval of non-rigid 3D shapes from multiple aspects. *Comput.-Aided Des.* **58**, 13–23 (2015)
4. Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J.: Image classification with the fisher vector: theory and practice. *Int. J. Comput. Vis.* **105**(3), 222–245 (2013)
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012)
6. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)* (2014)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
8. Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: Random erasing data augmentation. *arXiv preprint [arXiv:1708.04896](https://arxiv.org/abs/1708.04896)* (2017)
9. Wu, L., Wang, Y., Pan, S.: Exploiting attribute correlations: a novel trace lasso-based weakly supervised dictionary learning method. *IEEE Trans. Cybern.* **47**(12), 4497–4508 (2017)
10. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587 (2014)
11. Girshick, R.: Fast R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision* (2015)
12. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: *Advances in Neural Information Processing Systems*, pp. 91–99 (2015)
13. Zhong, Z., Lei, M., Cao, D., Fan, J., Li, S.: Class-specific object proposals re-ranking for object detection in automatic driving. *Neurocomputing* **242**, 187–194 (2017)

14. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3652–3661. IEEE (2017)
15. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camera style adaptation for person re-identification. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2018)
16. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: past, present and future. arXiv preprint [arXiv:1610.02984](https://arxiv.org/abs/1610.02984) (2016)
17. Wu, L., Wang, Y., Li, X., Gao, J.: What-and-where to match: deep spatially multiplicative integration networks for person re-identification. *Pattern Recognit.* **76**, 727–738 (2018)
18. Wu, L., Wang, Y., Gao, J., Li, X.: Deep adaptive feature embedding with local sample distributions for person re-identification. *Pattern Recognit.* **73**, 275–288 (2018)
19. Gopalan, R., Li, R., Chellappa, R.: Domain adaptation for object recognition: an unsupervised approach. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 999–1006. IEEE (2011)
20. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2066–2073. IEEE (2012)
21. Li, A., Shan, S., Chen, X., Gao, W.: Maximizing intra-individual correlations for face recognition across pose differences. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2009)
22. Hotelling, H.: Relations between two sets of variates. *Biometrika* **28**, 321–377 (1936)
23. Huang, K., Aviyente, S.: Sparse representation for signal classification. In: Advances in neural information processing systems, pp. 609–616 (2006)
24. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009)
25. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **15**(12), 3736–3745 (2006)
26. Olshausen, B.A., Field, D.J.: Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vis. Res.* **37**(23), 3311–3325 (1997)
27. Engan, K., Aase, S.O., Husoy, J.H.: Method of optimal directions for frame design. In: Acoustics, Speech, and Signal Processing, vol. 5, pp. 2443–2446. IEEE (1999)
28. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006)
29. Jia, Y., Salzmann, M., Darrell, T.: Factorized latent spaces with structured sparsity. In: Advances in Neural Information Processing Systems, pp. 982–990 (2010)
30. Qiu, Q., Patel, V.M., Turaga, P., Chellappa, R.: Domain adaptive dictionary learning. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012. LNCS, vol. 7575, pp. 631–645. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33765-9_45
31. Zheng, J., Jiang, Z., Phillips, P.J., Chellappa, R.: Cross-view action recognition via a transferable dictionary pair. In: BMVC (2012)
32. Shekhar, S., Patel, V.M., Nguyen, H.V., Chellappa, R.: Generalized domain-adaptive dictionaries. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 361–368. IEEE (2013)

33. Ni, J., Qiu, Q., Chellappa, R.: Subspace interpolation via dictionary learning for unsupervised domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 692–699. IEEE (2013)
34. Huang, D.A., Wang, Y.C.F.: Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2496–2503. IEEE (2013)
35. Zhu, F., Shao, L.: Enhancing action recognition by cross-domain dictionary learning. In: BMVC (2013)
36. Zhu, F., Shao, L.: Weakly-supervised cross-domain dictionary learning for visual recognition. *Int. J. Comput. Vis.* **109**(1–2), 42–59 (2014)
37. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset (2007)
38. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (SURF). *Comput. Vis. Image Underst.* **110**(3), 346–359 (2008)
39. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting visual category models to new domains. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6314, pp. 213–226. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15561-1_16