

# LEARNING GRASSMANN MANIFOLDS FOR OBJECT STATE DISCOVERY

Hao-Wei Lee

Department of Computer Science  
University of Toronto  
Toronto, Canada

Chia-Po Wei and Yu-Chiang Frank Wang

Research Center for IT Innovation  
Academia Sinica  
Taipei, Taiwan

## ABSTRACT

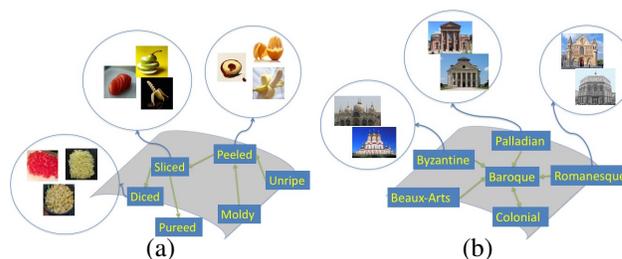
In this paper, we advocate the use of Grassmann manifolds for discovering object images in different states (e.g., unripe, peeled, etc.). We propose a novel dictionary learning algorithm, which derives the subspaces on a Grassmann manifold for describing each object state. By our introduced geodesic-flow constraint, our Grassmann manifold exhibits excellent capabilities in relating objects in distinct states (i.e., subspaces on the derived manifold), while the geodesics connecting different states can be viewed as transformations between the associated states. This is the reason why the use of our proposed Grassmann manifold can be applied to perform object state classification with improved performance. In our experiments, we provide quantitative and qualitative results to verify the effectiveness of our proposed method.

**Index Terms**— state recognition, classification, Grassmann manifold

## 1. INTRODUCTION

In the past few years, significant progress and improvements have been observed for visual classification (e.g., object recognition [1], scene recognition [2], or texture recognition [3]). However, when it comes to recognize different states of an observed object presented in an image, such problems have not yet been well studied. Take the fruit of apple for example (see Figure 1), a range of object states like unripe, sliced, or even moldy can be expected when seeing an image of apple. Since variations across different states are typically large even for the same type of object, the problem of *object state discovery* is a very challenging task.

Related to identifying different states of images, researchers proposed to explore the *attributes* of images [4, 5]. Without determining the relationships between object states, such information cannot be easily extended to object state discovery due to the lack of ability of performing state prediction. On the other hand, Xu et al. [6] presented a structured logistic regression model with latent variables, with the goal of classifying architecture styles of images. While relationships between architecture styles were considered, explicit



**Fig. 1:** Visualization of objects in different states on a Grassmann manifold. (a) states of vegetables and fruits. (b) style of architectures. For each state, its nearest neighbor is linked by an arrow.

definitions of deformable parts of textural/architectural regions were required, which makes the proposed method not applicable to general object style classification. Recently, Isola et al. [7] chose to discover object states and their relations in an image collection. Although promising performance was achieved, the relations between object state were predefined and cannot be automatically exploited from training.

Subspace learning is also a popular learning approach, which can be applied to visual classification tasks. For example, both Chang et al. [8] and Huang et al. [9] aimed at learning low-dimensional manifolds for describing face images with pose and illumination variations. As for action recognition, Slama et al. [10] and Turaga et al. [11] chose to model action images on a Grassmann manifold, and they showed that such manifolds exhibit sufficient capabilities in capturing image set variations.

Inspired by the above advances in subspace/manifold learning, we propose to address object state discovery by learning and exploring the associated Grassmann manifold. While recent domain adaptation approaches like [12, 13] have successfully considered such manifolds for recognizing cross-domain images (e.g., training and test images captured by different sensors), it is still a challenge to derive a proper Grassmann manifold for the purpose of object state discovery. As later detailed in Section 2, we propose a geodesic-flow constrained dictionary learning for learning Grassmann manifold, which can be applied for identifying and recognizing different states of objects.

The contributions of our paper can be summarized below:

- We are the first to explore the Grassmann manifold for identifying and recognizing different states of object images.
- We observe that subspaces derived on a Grassmann manifold reflect the states of each object, while their relationships can be properly modelled by geodesic flow information.
- We propose a geodesic-flow constrained dictionary learning algorithm, which allows us to exploit the relationships between dictionaries of different states for improved object state discovery.

## 2. OUR PROPOSED METHOD

### 2.1. Properties of Grassmann Manifold

For the completeness of discussions, we now provide a brief review of the Grassmann manifold. Denoted as  $\mathcal{G}(n, d)$ , a Grassmann manifold is a collection of all  $d$ -dimensional subspaces of  $\mathbb{R}^n$ , where  $n$  denotes the original feature dimension. A subspace can be viewed as a point on a Grassmann manifold, and each point on this manifold can be represented by an orthonormal basis  $\mathbf{P}$  (which does not need to be unique). If  $\text{span}(\mathbf{P}) = \text{span}(\mathbf{Q})$ , then  $\mathbf{P}$  and  $\mathbf{Q}$  correspond to the same point on a Grassmann manifold.

#### 2.1.1. Similarity Measure on a Grassmann Manifold

For each pair of points (i.e., subspaces) on a Grassmann manifold, we can calculate their similarity by their principal angles [14]. More specifically, the principal angle can be computed efficiently via singular value decomposition (SVD) of the inner product of two subspaces  $\mathbf{P}_1$  and  $\mathbf{P}_2$ , i.e.,

$$\mathbf{P}_1^T \mathbf{P}_2 = \mathbf{U}(\cos(\Theta))\mathbf{V}^T, \quad (1)$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are left and right-singular vectors of  $\mathbf{P}_1^T \mathbf{P}_2$ , respectively. The diagonal entries of  $\cos(\Theta)$  indicate the associated principal angles. Thus, the similarity between two subspaces on a Grassmann manifold is defined as  $\sum_{i=1}^d \cos^2(\theta_i)$  (recall that  $d$  is the dimension of the subspaces  $\mathbf{P}_1$  and  $\mathbf{P}_2$ ), which can be computed by  $\|\mathbf{P}_1^T \mathbf{P}_2\|_F^2$  [14].

#### 2.1.2. Geodesic Flow on a Grassmann Manifold

Geodesic describes the shortest path between two subspaces on a Grassmann manifold. The geodesic flow from subspaces  $P_1$  to  $P_2$  can be written in a parametric form [15] as follows:

$$\Psi(t) = \mathbf{Q} \exp(t\mathbf{B})\mathbf{Q}^T \mathbf{J}, \quad t \in [0, 1], \quad \Psi(0) = \mathbf{P}_1, \quad \Psi(1) = \mathbf{P}_2. \quad (2)$$

Note that  $\mathbf{Q} = [\mathbf{P}_1, \text{null}(\mathbf{P}_1^T)]$  is the orthogonal completion of  $\mathbf{P}_1$  that makes  $\mathbf{Q}^T \mathbf{P}_1 = \mathbf{J}$ ,  $\mathbf{J}$  indicates the first  $d$  columns of a  $n$ -by- $n$  identity matrix, and  $\mathbf{B}$  is a skew-symmetric matrix in the form of:

$$\begin{bmatrix} 0 & \mathbf{A}^T \\ -\mathbf{A} & 0 \end{bmatrix}, \quad \mathbf{A} \in \mathbb{R}^{(n-d) \times d}.$$

With the above definitions, we have  $\mathbf{P}_2 = \mathbf{W}\mathbf{P}_1$  and  $\mathbf{W} = \mathbf{Q} \exp(\mathbf{B})\mathbf{Q}^T$ . Thus, the matrix  $\mathbf{W}$  can be viewed as a transformation which projects  $\mathbf{P}_1$  to  $\mathbf{P}_2$  on the Grassmann manifold.

## 2.2. Object State Discovery

### 2.2.1. Grassmann Manifolds via Geodesic-Flow Constrained Dictionary Learning (GF-DL)

Assume that we observe data  $\mathbf{X}$  of  $K$  different states, i.e.,  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K]$  and  $\mathbf{X}_i \in \mathbb{R}^{n \times m}$ , where  $n$  denotes the original feature dimension and  $m$  is the number of instances of each state. When exploring the Grassmann manifold for describing the states of each object, we consider that each subspace (i.e., each point) on the constructed manifold corresponds to a particular state. Moreover, the relationship between different states would satisfy the similarity observed between the associated subspaces on this manifold.

Originally, principal components analysis (PCA) has been widely used to perform subspace learning for the observed image data subset (e.g.,  $\mathbf{X}_i$ ) [12]. In other words, the calculated eigenvectors of each  $\mathbf{X}_i$  represent the subspace on a Grassmann manifold. When it comes to exploring the Grassmann manifold for describing object states, one can directly construct  $K$  sets of eigenvectors from each  $\mathbf{X}_i$ . When recognizing the state of a test image, the reconstruction error of the projected test image in each state can be considered as the metric for performing state prediction.

However, the direct use of subspace learning techniques like PCA for constructing Grassmann manifolds fails to discover and preserve the relationship between images of different states. This is the reason why we propose a dictionary learning based algorithm, Geodesic-Flow constrained Dictionary Learning (GF-DL) for addressing the task of object state discovery. Inspired by recent success of dictionary learning approaches like [16], we aim at learning dictionary atoms for describing each state of the observed images, while the derived atoms of different states better describe the relative state information.

Recall that, given two subspaces on a Grassmann manifold, parametric form of geodesic between such two spans can be presented in terms of a transformation. That is, for two dictionaries  $\mathbf{D}_1$  and  $\mathbf{D}_2$ , there exists a transformation matrix  $\mathbf{W}$  satisfying  $\text{span}(\mathbf{W}\mathbf{D}_1) = \text{span}(\mathbf{D}_2)$ . In our proposed GF-DL, we aim at learning dictionaries representing each object state, while the geodesic between different states can be

---

**Algorithm 1** Geodesic-Flow Constrained Dictionary Learning (GF-DL) for Object State Discovery

---

**Require:** training data  $\mathbf{X}_1, \dots, \mathbf{X}_K$ ; parameter  $\eta$

- 1: initial  $\mathbf{d}_0$  by PCA;
  - 2: **while** not converge **do**
  - 3:   compute  $\mathbf{A}_i^*$  by OMP (4);
  - 4:   compute geodesic-flow constrained transformation matrix  $\mathbf{W}$ ;
  - 5:   update  $\mathbf{D}_{ref}^*$  by (9);
  - 6:   update  $\mathbf{D}_i^*$  by (10);
  - 7: **end while**
- 

properly preserved. Thus, our GF-DL solves the following optimization problem:

$$\begin{aligned} \min_{\mathbf{D}_i, \mathbf{A}_i, \mathbf{W}_i} & \sum_{i=1}^K \|\mathbf{X}_i - \mathbf{D}_i \mathbf{A}_i\|_F^2 - \eta \sum_{i \neq ref} \|(\mathbf{W}_i \mathbf{D}_{ref})^T \mathbf{D}_i\|_F^2 \\ \text{s.t.} & \|\mathbf{d}_i\|_2^2 = 1, \|\mathbf{a}_{(j)}\| \leq T_0, \mathbf{W}_i \in \text{Set}(\Psi), \end{aligned} \quad (3)$$

where  $\mathbf{X}_i$  denotes training image data of the  $i^{th}$  state,  $\mathbf{D}_i$  and  $\mathbf{A}_i$  are the associated dictionary and coefficients to be derived, respectively. The matrix  $\mathbf{W}_i$ , which relates  $\mathbf{D}_i$  and  $\mathbf{D}_{ref}$ , belongs to the set of transformation matrices as defined in (2). We note that, among  $K$  different states, our GF-DL randomly selects one state and takes the resulting dictionary as  $\mathbf{D}_{ref}$ . From (3), we see that the first term minimizes the reconstruction error between the observed image data and the derived low-dimensional representation, while the introduced second term enforces the geodesic flow constraint.

### 2.2.2. Optimization of GF-DL

It can be seen that, the objective function of (3) is not convex in terms of  $\mathbf{D}_i$ ,  $\mathbf{A}_i$ , and  $\mathbf{W}_i$ . However, it is convex when any two of the three variables are fixed. By applying the technique of alternative optimization (as did in [17]), we can learn  $\mathbf{D}_i$  and  $\mathbf{A}_i$ , and update the transformation matrix  $\mathbf{W}_i$  accordingly for preserving the relationships between different object states [13].

**Updating  $\mathbf{A}_i$ :** We first update each  $\mathbf{A}_i$  by fixing the dictionaries  $\mathbf{D}_i$  and the transformation matrix  $\mathbf{W}_i$ . The objective function shown below for updating  $\mathbf{A}_i$  can be easily solved by Orthogonal Matching Pursuit (OMP) [18]:

$$\begin{aligned} \min_{\mathbf{A}_i} & \|\mathbf{X}_i - \mathbf{D}_i \mathbf{A}_i\|_F^2 \\ \text{s.t.} & \|\mathbf{a}_{(j)}\| \leq T_0. \end{aligned} \quad (4)$$

**Updating  $\mathbf{W}_i$ :** When solving  $\mathbf{W}_i$ , we fix variables other than  $\mathbf{W}_i$ . Recall that the span of one subspace on a Grassmann

manifold can be transformed to another span by multiplying the transformation matrix  $\mathbf{W}$ . Based on  $\text{span}(\mathbf{D}_i) = \text{span}(\mathbf{W}_i \mathbf{D}_{ref})$  and (2), we have  $\mathbf{P}_2 = \mathbf{W} \mathbf{P}_1$  and  $\mathbf{W} = \mathbf{Q} \text{Exp}(\mathbf{B}) \mathbf{Q}^T$ . Let  $\mathbf{P}_1 = \mathbf{D}_{ref}$  and  $\mathbf{P}_2 = \mathbf{D}_i$ , we first derive orthogonal completion  $\mathbf{Q}$  of  $\mathbf{D}_{ref}$ , which is  $[\mathbf{D}_{ref}, \text{null}(\mathbf{D}_{ref}^T)]$ . To calculate  $\mathbf{B}$ , we compute CS decomposition of  $\mathbf{Q}^T \mathbf{D}_i$ :

$$\mathbf{Q}^T \mathbf{D}_i = \begin{bmatrix} \mathbf{V}_1 & 0 \\ 0 & \mathbf{V}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma} \\ -\mathbf{\Sigma} \end{bmatrix} \mathbf{V}^T. \quad (5)$$

The diagonal entries of  $\mathbf{\Gamma}$  and  $\mathbf{\Sigma}$  are the cosines and sines of principal angles between  $\mathbf{D}_{ref}$  and  $\mathbf{D}_i$ , and we obtain  $\mathbf{B} = \begin{bmatrix} 0 & \mathbf{A}^T \\ -\mathbf{A} & 0 \end{bmatrix}$ , where  $\mathbf{A} = \mathbf{V}_1 \sin^{-1}(\mathbf{\Sigma}) \mathbf{V}_2$ .

**Updating  $\mathbf{D}_{ref}$  and  $\mathbf{D}_i$ :** To update the reference dictionary  $\mathbf{D}_{ref}$ , we fix  $\mathbf{A}_i$ ,  $\mathbf{W}_i$  and the remaining  $\mathbf{D}_i$ . The objective function to be solved has the following form:

$$\begin{aligned} \min_{\mathbf{D}_{ref}} & \|\mathbf{X}_{ref} - \mathbf{D}_{ref} \mathbf{A}_{ref}\|_F^2 - \eta \sum_{i \neq ref} \|(\mathbf{W}_i \mathbf{D}_{ref})^T \mathbf{D}_i\|_F^2 \\ \text{s.t.} & \|\mathbf{d}_{ref}\|_2^2 = 1. \end{aligned} \quad (6)$$

We note that, the dictionary  $\mathbf{D}_{ref}$  is optimized by sequentially updating its columns. To solve the  $\ell^{th}$  column  $\mathbf{d}^\ell$ , we can simply ignore the irrelevant ones and rewrite the objective function as:

$$g(\mathbf{d}^\ell) = \left\| \mathbf{X}_{ref} - \sum_{j \neq \ell} d^j \mathbf{a}_{(j)} - \mathbf{d}^\ell \mathbf{a}_{(\ell)} \right\|_F^2 - \eta \sum_{i \neq ref} \|\mathbf{D}_i^T \mathbf{W}_i \mathbf{d}^\ell\|_F^2. \quad (7)$$

Let  $\hat{\mathbf{X}} = \mathbf{X}_{ref} - \sum_{j \neq \ell} d^j \mathbf{a}_{(j)}$ , the objective function can be further simplified as:

$$g(\mathbf{d}^\ell) = \left\| \hat{\mathbf{X}} - \mathbf{d}^\ell \mathbf{a}_{(\ell)} \right\|_F^2 - \eta \sum_{i \neq ref} \|\mathbf{D}_i^T \mathbf{W}_i \mathbf{d}^\ell\|_F^2. \quad (8)$$

Now, we can calculate the optimal  $\mathbf{d}^\ell$  by letting  $\frac{\partial g(\mathbf{d}^\ell)}{\partial \mathbf{d}^\ell} = 0$ , i.e.,

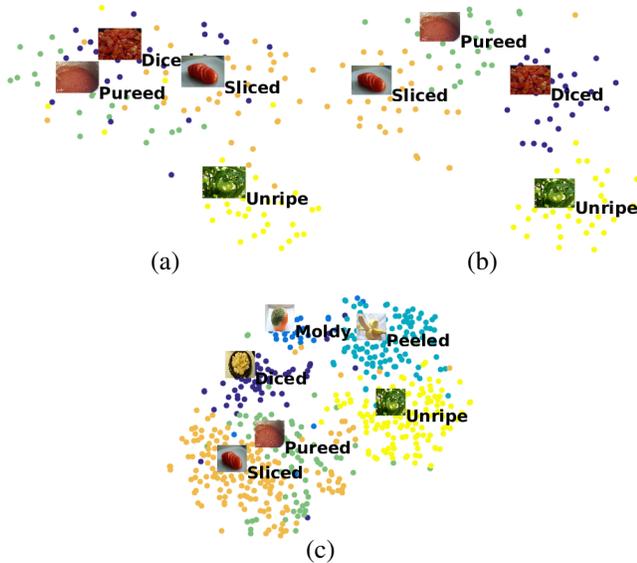
$$\mathbf{d}^{\ell*} = \left( \|\mathbf{a}_{(\ell)}\|_2^2 - \eta \sum_{i \neq ref} \mathbf{W}_i^T \mathbf{D}_i \mathbf{D}_i^T \mathbf{W}_i \right)^{-1} \hat{\mathbf{X}} \mathbf{a}_{(\ell)}^T. \quad (9)$$

Since the column vectors of each dictionary need to be of unit length, we normalize the dictionary after the above derivation, i.e.,  $\mathbf{d}^\ell = \mathbf{d}^\ell / \|\mathbf{d}^\ell\|$ . We also require the re-scaling of the corresponding  $\mathbf{a}_{(\ell)}$  by  $\mathbf{a}_{(\ell)} = \|\mathbf{d}^\ell\| \mathbf{a}_{(\ell)}$ .

Similar to the update of  $\mathbf{D}_{ref}$ , we calculate  $\mathbf{D}_i$  with the other variables fixed. The optimal  $\mathbf{d}^\ell$  can be derived by letting  $\frac{\partial g(\mathbf{d}^\ell)}{\partial \mathbf{d}^\ell} = 0$ . With the normalization of  $\mathbf{d}^\ell$  and re-scaling of  $\mathbf{a}_{(\ell)}$ , we have:

$$\mathbf{d}^{\ell*} = \left( \|\mathbf{a}_{(\ell)}\|_2^2 - \eta \mathbf{W}_i \mathbf{D}_{ref} \mathbf{D}_{ref}^T \mathbf{W}_i^T \right)^{-1} \hat{\mathbf{X}} \mathbf{a}_{(\ell)}^T. \quad (10)$$

The pseudo code of our proposed method is summarized in Algorithm 1.



**Fig. 2:** Visualization of object images in different states via t-SNE. While (a) indicates the direct use of CNN features for describing object images in different states, the use of our derived Grassmann manifold for identifying tomatoes and all object categories in different states are shown in (b) and (c), respectively.

### 3. EXPERIMENTS

#### 3.1. State Classification of Fruits and Vegetables

We now conduct our first experiments on object state classification using the dataset provided in [19]. We choose object images in 6 states (i.e., diced, moldy, peeled, pureed, sliced, and unripe) from 10 objects: apple, banana, berry, fruit, lemon, orange, pear, persimmon, potato, tomato, and vegetable.

To describe each object image, we apply Caffe [20] to pre-train a convolutional neural network (CNN) on the ImageNet database, and use CNN as the feature extractor. We select object images in different states from the categories of apple and lemon as the training set, and those in the remaining categories as the test images. For our GF-DL algorithm, the dictionary size is set as 20 for every state with  $T_0 = 5$  and  $\eta = 0.4$ . To compare the classification performance (using CNN as the feature extractor), we consider linear SVM [21], PCA, K-SVD [19], Label Consistent K-SVD (LC-KSVD) [22], and Fisher Discrimination Dictionary Learning (FDDL) [23], in which SVM can be considered as the baseline method.

Table 1 lists the recognition rates of different methods. It can be seen from Table 1 that dictionary-learning based methods (PCA, K-SVD, LC-KSVD, FDDL, and our GF-DL) deliver better classification performance than SVM. Among all test methods, our GF-DL achieves the best classification performance. We note that both LC-KSVD and FDDL utilize additional label information for dictionary learning, but their recognition rates are no better than that of K-SVD. This is be-

**Table 1:** Classification performance of different approaches.

SVM	PCA	K-SVD	LC-KSVD	FDDL	<b>GF-DL</b>
44.92	51.97	55.8	54.5	48.8	<b>57.35</b>

**Table 2:** Classification performance of different approaches on Architectural Style.

SVM	PCA	K-SVD	LC-KSVD	FDDL	<b>GF-DL</b>
66.8	71.5	65.6	66.0	61.6	<b>73.1</b>

cause they do not exploit the relationships between different states.

In addition to quantitative evaluation, we further visualize the projected data in a lower-dimensional space. Figure 2 illustrates the projection of object images with different states in a 2D subspace by t-SNE [24]. We see that the use of our proposed GF-DL results in improved state relationships in the 2D subspace as shown in Figures 2b and 2c. We also measure the similarities between different object states (i.e., subspaces) on our Grassmann manifold. Based on the metric of nearest neighbors, we plot the relationships between different object states in Figure 1 (i.e., 1-NN graph). Figure 1a depicts the transformation path between different states (e.g., unripe→peeled→sliced→diced). The above quantitative and qualitative results support the effectiveness of our GF-DL for object state discovery.

#### 3.2. Style Classification of Architectures

The second experiment is conducted on 6 selected styles from the Architectural Style dataset [6], in which the 6 architectural styles are Baroque, Beaux-Arts, Byzantine, Colonial, Palladian, and Romanesque. This relatively simple problem aims to classify states (architectural style) of similar objects (buildings). The recognition rates of different methods are shown in Table 2. From Table 2, our method still achieves the best classification performance compared to other test methods. In addition, our constructed dictionaries can also be used to establish the relationships between architectural styles as shown in Figure 1b.

### 4. CONCLUSIONS

In this paper, we proposed a novel dictionary learning algorithm to construct Grassmann manifolds, which can be applied for solving the task of object state discovery. Due to the introduced geodesic-flow constraint, we not only describe each object state as a subspace on the constructed Grassmann manifold, the relationship between different states can be properly preserved. This is the reason why we can apply our approach for recognizing different states of object images. Finally, we provided experimental results on real-world object images in different states. From both quantitative and qualitative results, the use of our approach for object state discovery can be successfully verified.

## References

- [1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and F.-F. Li, "ImageNet Large Scale Visual Recognition Challenge," *IJCV*, 2015.
- [2] B.-L. Zhou, A. Lapedriza, J.-X. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *NIPS*, 2014.
- [3] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," in *IEEE, CVPR*, 2015.
- [4] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth, "Describing objects by their attributes," in *IEEE, CVPR*, 2009.
- [5] D. Parikh and K. Grauman, "Relative attributes," in *IEEE, ICCV*, 2011.
- [6] Z. Xu, D.-C. Tao, Y. Zhang, J.-J. Wu, and A.-C. Tsoi, "Architectural style classification using multinomial latent logistic regression," in *ECCV*, 2014.
- [7] P. Isola, J. J. Lim, and E. H. Adelson, "Discovering states and transformations in image collections," in *IEEE, CVPR*, 2015.
- [8] J.-M. Chang, M. Kirby, and C. Peterson, "Set-to-set face recognition under variations in pose and illumination," in *Proc. Biometrics Symposium*, 2007.
- [9] Z.-W. Huang, R.-P. Wang, S.-G. Shan, and X.-L. Chen, "Projection metric learning on Grassmann manifold with application to video based face recognition," in *IEEE, CVPR*, 2015.
- [10] R. Slama, H. Wannous, M. Daoudi, and A. Srivastava, "Accurate 3D action recognition using learning on the Grassmann manifold," *Pattern Recognition*, 2015.
- [11] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa, "Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition," *IEEE TPAMI*, 2011.
- [12] B.-Q. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *IEEE, CVPR*, 2012.
- [13] R. Gopalan, R.-N. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *IEEE, ICCV*, 2011.
- [14] M. T. Harandi, M. Salzmann, S. Jayasumana, R. Hartley, and H.-D. Li, "Expanding the family of Grassmannian kernels: An embedding perspective," in *ECCV*, 2014.
- [15] K. A. Gallivan, A. Srivastava, X.-W. Liu, and P. Van Dooren, "Efficient algorithms for inferences on Grassmann manifolds," in *IEEE WSSP*, 2003.
- [16] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *IEEE, CVPR*, 2010.
- [17] S. Kong and D.-H. Wang, "A dictionary learning approach for classification: separating the particularity and the commonality," in *ECCV*, 2012.
- [18] S. G. Mallat and Z.-F. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE TSP*.
- [19] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE TSP*.
- [20] Y.-Q. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *ACM MM*, 2014.
- [21] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM TIST*, 2011.
- [22] Z.-L. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," in *IEEE, CVPR*, 2011.
- [23] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *IEEE ICCV*, 2011.
- [24] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *JMLR*, 2008.