Submodular Dictionary Learning for Sparse Coding

Zhuolin Jiang   Guangxiao Zhang   Larry S. Davis

University of Maryland at College Park, MD, 20742

1. Overview

- **Goal**
  - Present a supervised algorithm for *efficiently* learning a compact and discriminative dictionary for sparse representation.
- **Approach**
  - A dataset is mapped into an undirected k-nearest neighbor graph $G=(V, E)$. The discriminative dictionary learning is modeled as a graph topology selection problem.
  - A *monotonic* and *submodular* objective function for dictionary learning consists of two terms: the entropy rate of a random walk on a graph and a discriminative term.
  - The objective function is optimized by a highly efficient greedy algorithm by using the submodularity and monotonic increasing properties of the objective function and the *matroid* constraint.
  - This simple greedy algorithm gives a near-optimal solution with a (1/2)-approximation bound [5].

2. Related Work

- Sparse Coding has been successfully applied to a variety of problems in computer vision such as face recognition [1]. The SRC algorithm [1] employs the entire set of training samples to form a dictionary.
- K-SVD [2]: Efficiently learn an over-complete dictionary with a small size. It focuses on representational power, but it does not consider discrimination.
- Discriminative dictionary learning approaches:
  - ❑ Constructing a separate dictionary for each class.
  - ❑ Adding discriminative terms into the objective function of dictionary learning [3].
- The diminishing return property of a submodular function has been employed in applications such as sensor placement, clustering and superpixel segmentation [4].

3. Preliminaries

- **Submodularity**

  Let E be a finite set. A set function $F: 2^E \rightarrow R$ is submodular if

  $$F(A \cup \{a_1\}) - F(A) \geq F(A \cup \{a_1, a_2\}) - F(A \cup \{a_2\})$$

  for all $A \subseteq E$ and $a_1, a_2 \in E \setminus A$. *(diminishing returns property)*
- **Matroid**

  Let E be a finite set and $\mathcal{I}$ a collection of subsets of E. A matroid is an ordered pair $\mathcal{M} = (E, \mathcal{I})$ satisfying three conditions:

  (a) $\emptyset \in \mathcal{I}$; (b) if $A \subseteq B$ and $B \in \mathcal{I}$, then $A \in \mathcal{I}$; (c) if $A \in \mathcal{I}, B \in \mathcal{I}$ and $|A| < |B|$, then there is an element $x \in B - A$ such that $A \cup x \in \mathcal{I}$.

4. Submodular Dictionary Learning

- **Monotonic and Submodular Objective Function**

  ❑ Consists of an entropy rate term $\mathcal{H}(A)$ and a discriminative term $\mathcal{Q}(A)$:

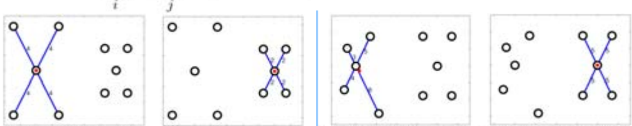
  $$\max_A \mathcal{F}(A) = \mathcal{H}(A) + \lambda \mathcal{Q}(A) \ s.t. \ A \subseteq E \ and \ N_A \geq K,$$

  A: selected subset of edge set E; N_A: number of connected components induced by A

- **Entropy Rate of a Random Walk**

$$\mathcal{H}(A) = -\sum_i \mu_i \sum_j P_{i,j}(A) \log P_{i,j}(A)$$

μ_i : Stationary probability of vertex v_i
$P_{i,j}$: Transition probability from v_i to v_j

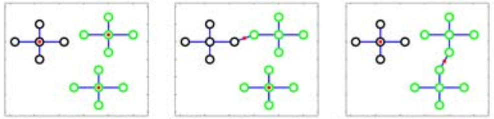


(a) Entropy Rate = 0.03 (b) Entropy Rate = 0.43 (c) Entropy Rate = 0.22 (d) Entropy Rate = 0.24

- **Discriminative Term**

$$\mathcal{Q}(A) = \frac{1}{C} \sum_{i=1}^{N_A} \max_y N_y^i - N_A$$

N_y^i : Number of elements from class y in cluster i



(a) Disc. Fun. = −2.00 (b) Disc. Fun. = −1.33 (c) Disc. Fun. = −1.00

- **Optimization**

  ❑ The cycle free constraint and connected component constraint, $N_A \geq K$, induces a matroid $\mathcal{M} = (E, \mathcal{I})$. Dictionary learning is achieved via maximizing a submodular function subject to a matroid constraint: $\max_A \mathcal{F}(A) \ s.t. \ A \in \mathcal{I}$

Algorithm 1 Submodular Dictionary Learning (SDL)

Input: $G = (V, E), w, K, \lambda$ and $\mathcal{N}$
Output: D
Initialization: $A \leftarrow \emptyset, D \leftarrow \emptyset$
for $N_A > K$ **do**
  $\tilde{e} = \underset{A \cup \{e\} \in \mathcal{I}}{\operatorname{argmax}} \mathcal{F}(A \cup \{e\}) - \mathcal{F}(A)$
  $A \leftarrow A \cup \{\tilde{e}\}$
end for
for each subgraph S_i in $G = (V, A)$ **do**
  $D \leftarrow D \cup \{\frac{1}{|S_i|} \sum_{j: v_j \in S_i} v_j\}$
end for

- **Classification**

  ❑ **Face and Object recognition**

  For a test image y_i, first compute its sparse representation:

  $$z_i = \arg\min_{z_i} \|y_i - Dz_i\|_2^2 \ s.t. \ \|z_i\|_0 \leq s$$

  Then the label of y_i is the index i corresponding to the largest element of a class label vector $l = Wz_i$.

  ❑ **Action Classification**

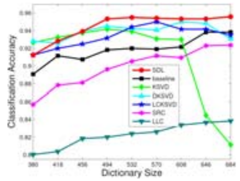
  First compute a sparse representations for each frame, then employ dynamic time warping to align two sequences in the sparse representation domain; next a K-NN classifier is used for recognition.

5. Experiments

- **Experimental Setup**
  - ❑ **Random face-based features**
    - dims: 504 (Extended Yale)
  - ❑ **Spatial pyramid features**
    - 1024 bases
    - dims: 3000 (Caltech101)
  - ❑ **Joint Shape and Motion features**
    - dims: 512 (Keck Gesture)

- **Extended Yale**

  ❑ Classification accuracy comparison



❑ Computation time (s) for dictionary training

Dict. size	418	456	494	532	570	608	646	684
SDL	0.9	1.0	0.9	0.9	0.9	1.0	0.9	0.9
K-SVD [1]	52.6	56.1	59.8	64.9	67.9	72.2	76.2	78.0
D-KSVD [35]	53.1	56.9	60.5	65.8	68.1	74.9	77.6	79.2
LC-KSVD [12]	67.2	72.6	78.3	86.5	90.7	97.8	104.4	112.3

- **Caltech101**

  ❑ Classification accuracy comparison

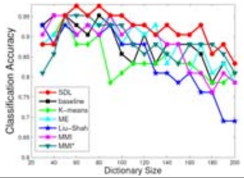
Training Images	5	10	15	20	25	30
Malik [34]	46.6	55.8	59.1	62.0	-	66.20
Lazebnik [15]	-	-	56.4	-	-	64.6
Griffin [9]	44.2	54.5	59.0	63.3	65.8	67.60
Irani [2]	-	-	65.0	-	-	70.40
Grauman [11]	-	-	61.0	-	-	69.10
Venkatesh [25]	-	-	42.0	-	-	-
Gemert [7]	-	-	-	-	-	64.16
Yang [31]	-	-	67.0	-	-	73.20
Wang [29]	51.15	59.77	65.43	67.74	70.16	73.44
SRC [30]	48.8	60.1	64.9	67.7	69.2	70.7
K-SVD [1]	49.8	59.8	65.2	68.7	71.0	73.2
D-KSVD [35]	49.6	59.5	65.1	68.6	71.1	73.0
LC-KSVD [12]	54.0	63.1	**67.7**	70.5	72.3	73.6
SDL	**55.3** ±0.5	**63.4** ± 0.5	67.5 ± 0.3	**70.7** ± 0.3	**73.1** ± 0.4	**75.3** ± 0.4

❑ Computation time (s) for dictionary training

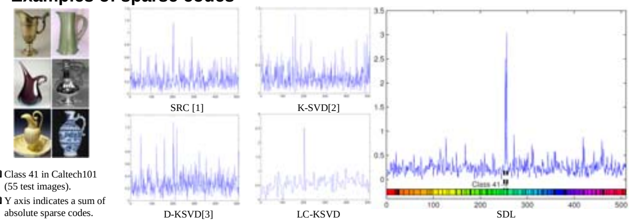
Dict. size	306	510	714	918	1122	1326	1530
SDL	37.5	36.7	36.6	36.9	37.1	36.7	36.7
K-SVD [1]	578.3	790.1	1055	1337	1665	2110	2467
D-KSVD [35]	560.1	801.3	1061	1355	1696	2081	2551
LC-KSVD [12]	612.1	880.6	1182	1543	1971	2496	3112

- **Keck Gesture Dataset**

  ❑ Classification accuracy comparison



- **Examples of sparse codes**



6. Key References

1. J. Wright, A. Yang, A. Ganesh, S. Sastry and Y. Ma. **Robust face recognition via sparse representation**, TPAMI 2009.
2. M. Aharon, M. Elad and A. Bruchstein. **K-SVD: An algorithm for designing over-complete dictionaries for sparse representation**. Sig. Proc., 2006.
3. Q. Zhang and B. Li. **Discriminative k-svd for dictionary learning in face recognition**, CVPR 2010.
4. M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa. **Entropy rate superpixel segmentation**, CVPR 2011.
5. G. Nemhauser, L. Wolsey, and M. Fisher. **An analysis of the approximations for maximizing submodular set functions**. Mathematical Programming, 1978