



# Sparse Dictionary-based Representation and Recognition of Action Attributes

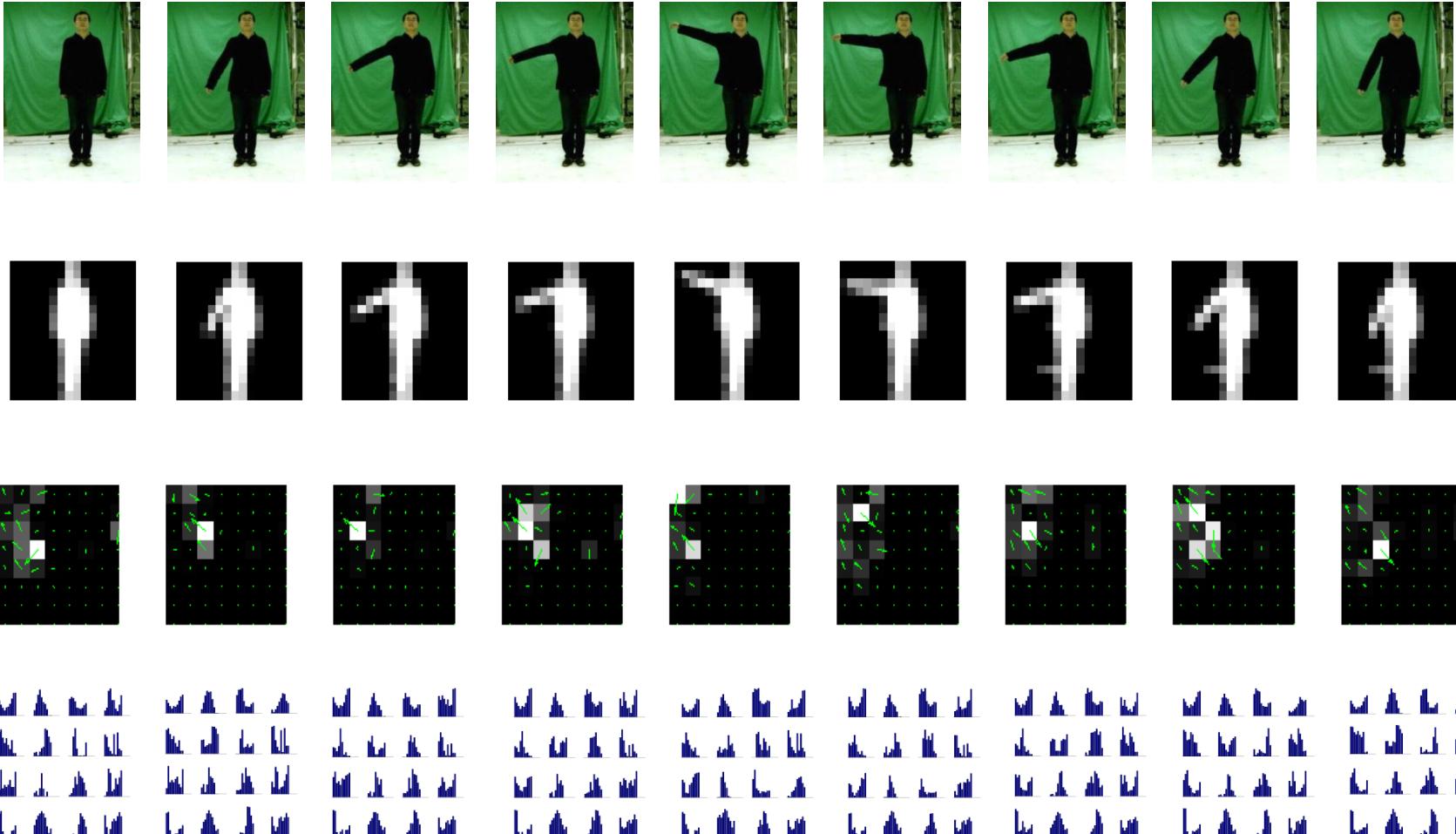
---

**Qiang Qiu, Zhuolin Jiang, Rama Chellappa**

Center for Automation Research,  
Institute for Advanced Computer Studies  
University of Maryland, College Park

[qiu@cs.umd.edu](mailto:qiu@cs.umd.edu), [{zhuolin, rama}@umiacs.umd.edu](mailto:{zhuolin, rama}@umiacs.umd.edu)

# Action Feature Representation



# Action Sparse Representation

Sparse code

Action

Dictionary

0.43	0
0.63	0
0	0.64
0	0.53
-0.33	-0.40
0	0.35
-0.36	0
0	0
0	0

$$\left( \begin{array}{c} \text{Action} \\ \uparrow \\ \text{Y} \end{array} \right) = \left( \begin{array}{c} \text{Dictionary} \\ \text{Y} \\ \text{Y} \end{array} \right)$$

$$\text{Y} = 0.43 \times \text{Y}_1 + 0.63 \times \text{Y}_2 - 0.33 \times \text{Y}_3 - 0.36 \times \text{Y}_4$$

$$\text{Y} = 0.64 \times \text{Y}_1 + 0.53 \times \text{Y}_2 - 0.40 \times \text{Y}_3 + 0.35 \times \text{Y}_4$$

# K-SVD



## Sparse codes

### Input signals

$$\left[ \begin{array}{c} y_1 \\ y_2 \end{array} \right] = \left[ \begin{array}{cccc} d_1 & d_2 & d_3 & \dots \end{array} \right] \left[ \begin{array}{c} \text{Y} \\ \text{D} \end{array} \right]$$

The input signals  $y_1$  and  $y_2$  are represented as 2x1 vectors. The dictionary  $D$  is a matrix where each column  $d_i$  is a 2x1 vector representing a signal. The product  $DY$  results in a 2x10 matrix where each row is a sparse code vector  $x_i$ .

### Dictionary

$\mathbf{x}_1$	$\mathbf{x}_2$
0.43	0
0.63	0
0	0.64
0	0.53
-0.33	-0.40
0	0.35
-0.36	0
0	0
0	0

The sparse codes  $\mathbf{x}$  are represented as a 10x2 matrix. The first column contains the values for  $\mathbf{x}_1$  and the second column contains the values for  $\mathbf{x}_2$ . The entries are rounded to two decimal places.

$$\arg \min_{D, X} \| Y - DX \|^2 \quad s.t. \quad \forall i, \| x_i \|_0 \leq T$$

### ■ K-SVD [1]

- Input: signals  $Y$ , dictionary size, sparsity  $T$
- Output: dictionary  $D$ , sparse codes  $X$

---

[1] M. Aharon and M. Elad and A. Bruckstein, K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation, IEEE Trans. on Signal Process, 2006<sup>4</sup>

# Objective



Learn a  
**Compact** and **Discriminative**  
Dictionary.

# Probabilistic Model for Sparse Representation



- A Gaussian Process
- Dictionary Class Distribution

# More Views of Sparse Representation

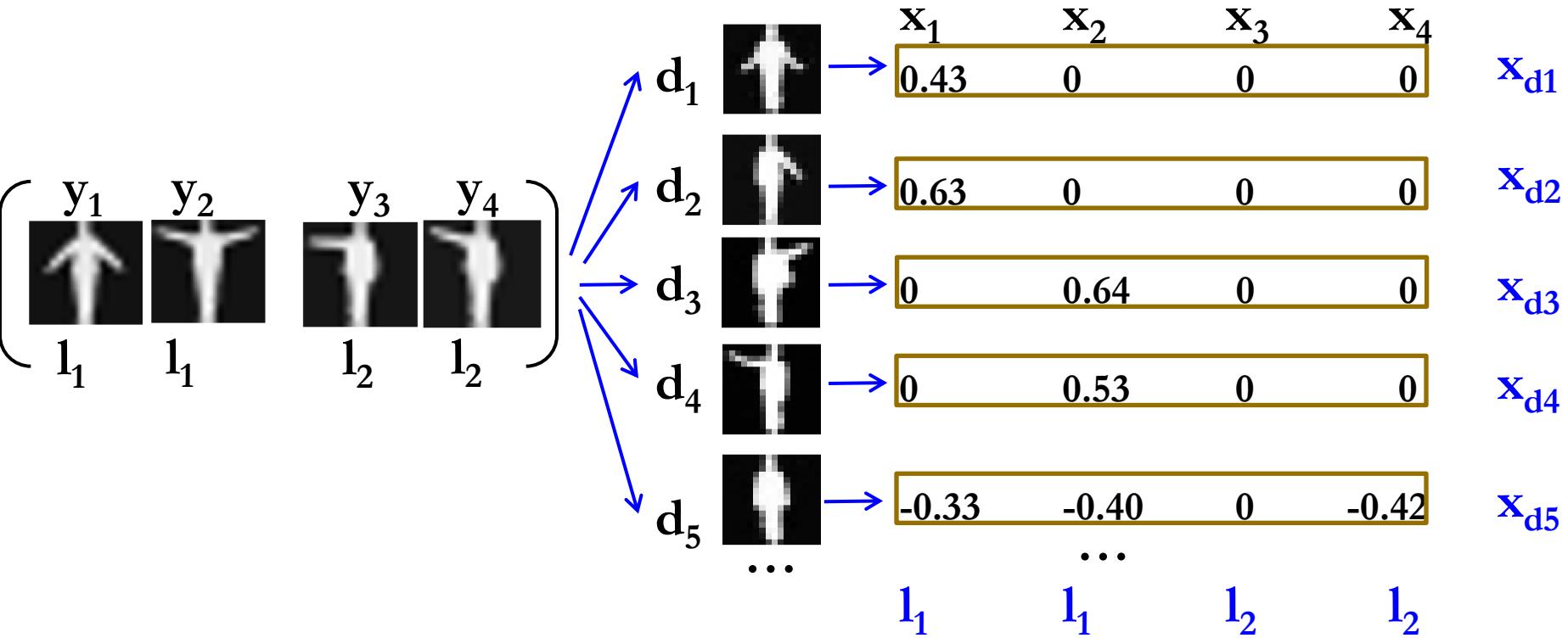


$$\left( \begin{array}{cc} y_1 & y_2 \\ l_1 & l_1 \\ \hline y_3 & y_4 \\ l_2 & l_2 \end{array} \right) = \mathbf{x}_{d1} \left( \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \\ l_1 & l_1 & l_2 & l_2 \end{array} \right)$$

The diagram illustrates sparse representation. On the left, four input vectors  $y_1, y_2, y_3, y_4$  are shown above their corresponding sparse representations  $l_1, l_1, l_2, l_2$ . A red circle highlights the first element  $d_1$  of the dictionary  $\mathbf{x}_{d1}$ , which corresponds to the first row of the matrix. The matrix  $\mathbf{x}_{d1}$  is defined by columns  $x_1, x_2, x_3, x_4$  and rows  $l_1, l_1, l_2, l_2$ . The first row of the matrix is circled in red.

$x_1$	$x_2$	$x_3$	$x_4$
0.43	0	0	0
0.63	0	0	0
0	0.64	0	0
0	0.53	0	0
-0.33	-0.40	0	-0.42
0	0.35	0	0
-0.36	0	0	0
0	0	0	0
0	0	-0.28	0
0	0	0.698	0.42
0	0	0.37	0.47
0	0	0.25	0
0	0	0	0.32

# A Gaussian Process

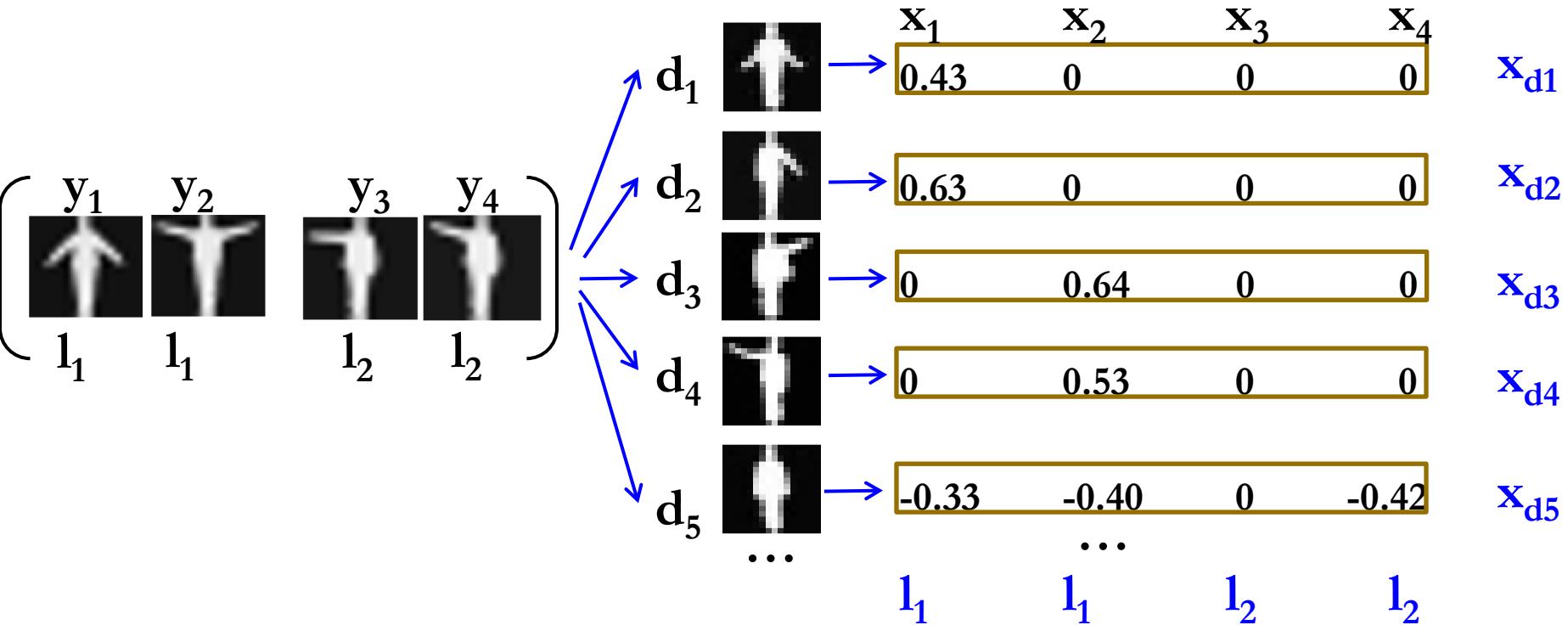


## A Gaussian Process

- Covariance function entry:  $K(i,j) = \text{cov}(x_{di}, x_{dj})$
- $P(X_{d^*} | X_{D^*})$  is a Gaussian with a closed-form conditional variance

$$\mathbb{V}(d^* | D^*) = \mathcal{K}_{(d^*, d^*)} - \mathcal{K}_{(d^*, D^*)}^T \mathcal{K}_{(D^*, D^*)}^{-1} \mathcal{K}_{(d^*, D^*)}$$

# Dictionary Class Distribution



## Dictionary Class Distribution

- $P(L | d_i)$ ,  $L \in [1, M]$ 
  - aggregate  $|x_{di}|$  based on class labels to obtain a  $M$  sized vector
    - $P(L=l_1 | d_5) = (0.33+0.40)/(0.33+0.40+0.42) = 0.6348$
    - $P(L=l_2 | d_5) = (0+0.42)/(0.33+0.40+0.42) = 0.37$

# Dictionary Learning Approaches



- Maximization of Joint Entropy (ME)
- Maximization of Mutual Information (MMI)
  - Unsupervised Learning (MMI-1)
  - Supervised Learning (MMI-2)

# Maximization of Joint Entropy (ME)

- Initialize dictionary using k-SVD

$$D^o = \left( \begin{array}{c|c|c|c|c|c|c|c|c|c|c} \text{Image 1} & \text{Image 2} & \text{Image 3} & \text{Image 4} & \text{Image 5} & \text{Image 6} & \text{Image 7} & \text{Image 8} & \text{Image 9} & \text{Image 10} \\ \text{Image 11} & \text{Image 12} & \text{Image 13} & \text{Image 14} & \text{Image 15} & \text{Image 16} & \text{Image 17} & \text{Image 18} & \text{Image 19} & \text{Image 20} \\ \text{Image 21} & \text{Image 22} & \text{Image 23} & \text{Image 24} & \text{Image 25} & \text{Image 26} & \text{Image 27} & \text{Image 28} & \text{Image 29} & \text{Image 30} \\ \text{Image 31} & \text{Image 32} & \text{Image 33} & \text{Image 34} & \text{Image 35} & \text{Image 36} & \text{Image 37} & \text{Image 38} & \text{Image 39} & \text{Image 40} \end{array} \right)$$

- Start with  $D^* = \phi$
- Until  $|D^*| = k$ , iteratively choose  $d^*$  from  $D^o \setminus D^*$ ,

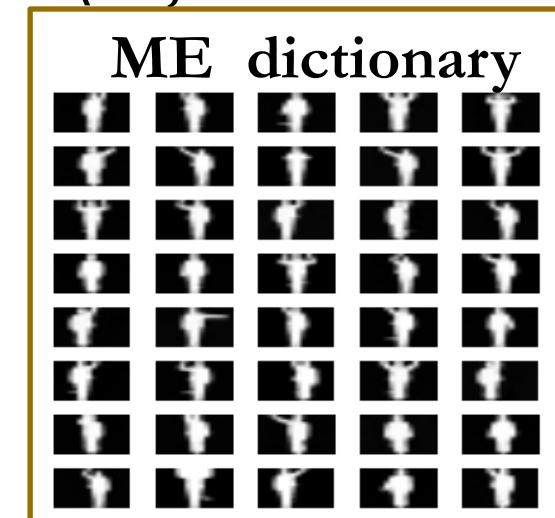
$$d^* = \arg \max_d H(d | D^*)$$

Where

$$H(d^* | D^*) = \frac{1}{2} \log(2\pi e V(d^* | D^*))$$

- A good approximation to ME criteria

$$\arg \max_D H(D)$$



---


$$\begin{aligned}
 \text{Image 1} &\parallel \text{Image 2} = \text{Image 1} + \text{Image 2} + \text{Image 3} + \text{Image 4} \\
 \text{Image 1} &\parallel \text{Image 2} = \text{Image 1} + \text{Image 2} + \text{Image 3} + \text{Image 4}
 \end{aligned}$$

# Maximization of Mutual Information for Unsupervised Learning (MMI-1)



- Initialize dictionary using k-SVD

$$D^o = \left\{ \begin{array}{c} \text{matrix of images} \\ \text{rows: images} \\ \text{columns: features} \end{array} \right\}$$

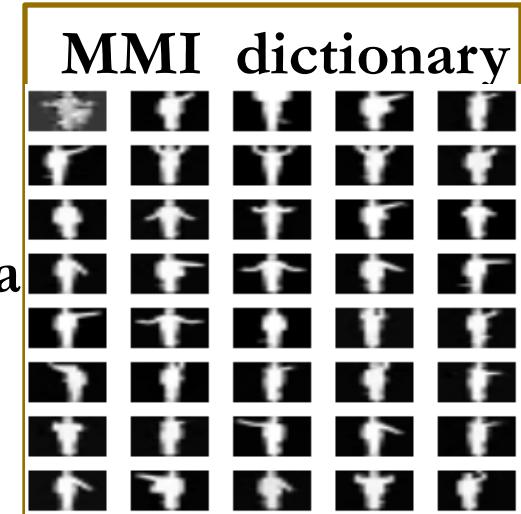
- Start with  $D^* = \emptyset$
- Until  $|D^*| = k$ , iteratively choose  $d^*$  from  $D^o \setminus D^*$ ,

$$d^* = \arg \max_d H(d | D^*) - H(d | D^o \setminus (D^* \cup d))$$

- A near-optimal approximation to MMI criteria

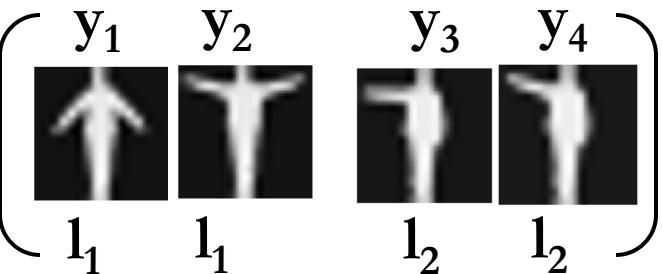
$$\arg \max_D I(D; D^o \setminus D)$$

Within  $(1-1/e)$  of the optimum



$$\begin{aligned} \text{original image} &= \text{image 1} + \text{image 2} + \text{image 3} + \text{image 4} \\ \text{reconstructed image} &= \text{image 1} + \text{image 2} + \text{image 3} + \text{image 4} \end{aligned}$$

# Revisit



	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	
$d_1$	0.43	0	0	0	$\mathbf{x}_{d1}$
$d_2$	0.63	0	0	0	$\mathbf{x}_{d2}$
$d_3$	0	0.64	0	0	$\mathbf{x}_{d3}$
$d_4$	0	0.53	0	0	$\mathbf{x}_{d4}$
$d_5$	-0.33	-0.40	0	-0.42	$\mathbf{x}_{d5}$
	...				
	$l_1$	$l_1$	$l_2$	$l_2$	

## Dictionary Class Distribution

- $P(L|d_i)$ ,  $L \in [1, M]$ 
  - aggregate  $|x_{di}|$  based on class labels to obtain a  $M$  sized vector
    - $P(l_1|d_5) = (0.33+0.40)/(0.33+0.40+0.42) = 0.6348$
    - $P(l_2|d_5) = (0+0.42)/(0.33+0.40+0.42) = 0.37$
- $P(L_d) = P(L/d)$
- $P(L_D) = P(L/D)$ , where  $P(L|D^*) = \frac{1}{|D^*|} \sum_{d_i \in D^*} P(L|d_i)$

# Maximization of Mutual Information for Supervised Learning (MMI-2)



- Initialize dictionary using k-SVD

$$D^o = \left[ \begin{array}{cccccccccc} \text{image 1} & \text{image 2} & \text{image 3} & \text{image 4} & \text{image 5} & \text{image 6} & \text{image 7} & \text{image 8} & \text{image 9} & \text{image 10} \\ \text{image 11} & \text{image 12} & \text{image 13} & \text{image 14} & \text{image 15} & \text{image 16} & \text{image 17} & \text{image 18} & \text{image 19} & \text{image 20} \\ \text{image 21} & \text{image 22} & \text{image 23} & \text{image 24} & \text{image 25} & \text{image 26} & \text{image 27} & \text{image 28} & \text{image 29} & \text{image 30} \\ \text{image 31} & \text{image 32} & \text{image 33} & \text{image 34} & \text{image 35} & \text{image 36} & \text{image 37} & \text{image 38} & \text{image 39} & \text{image 40} \end{array} \right]$$

- Start with  $D^* = \emptyset$
- Until  $|D^*| = k$ , iteratively choose  $d^*$  from  $D^o \setminus D^*$ ,

$$\begin{aligned} d^* = \arg \max_d [H(d | D^*) - H(d | D^o \setminus (D^* \cup d))] \\ + \lambda [H(L_d | L_{D^*}) - H(L_d | L_{D^o \setminus (D^* \cup d)})] \end{aligned}$$

- MMI-1 is a special case of MMI-2 with  $\lambda=0$ .

# Other learning methods

- K-means
- Liu-shah [1]

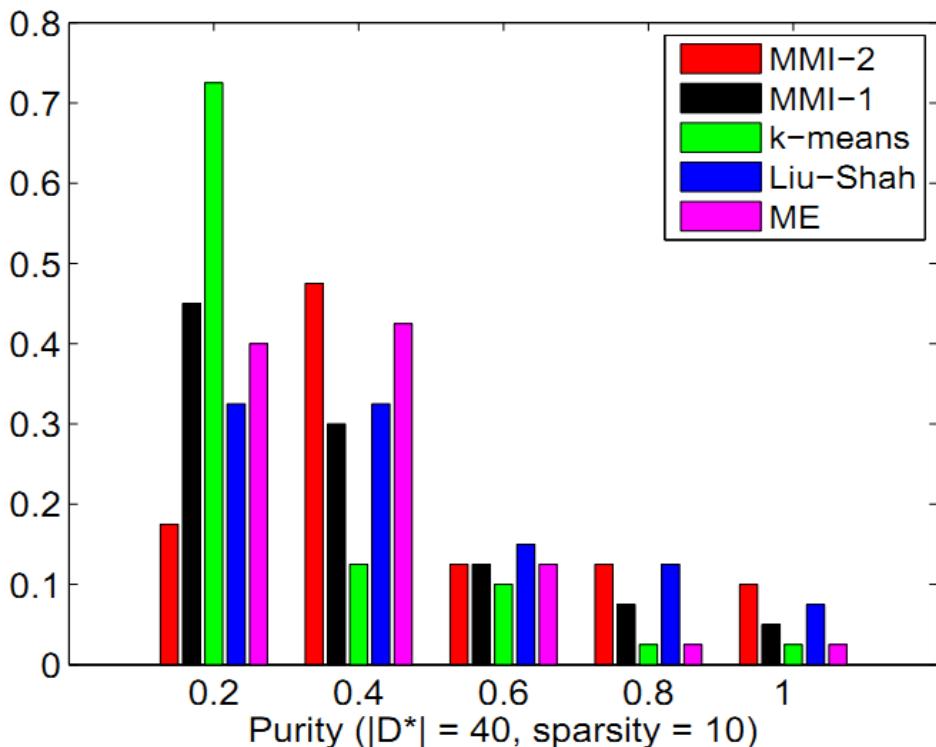
$$\begin{aligned} \Delta I(d_1, d_2) = & \sum_{L \in [1, M], i=1,2} p(d_i) p(L|d_i) \log p(L|d_i) \\ & - p(d_i) p(L|d_i) \log p(L|d^*) \end{aligned}$$

**where**

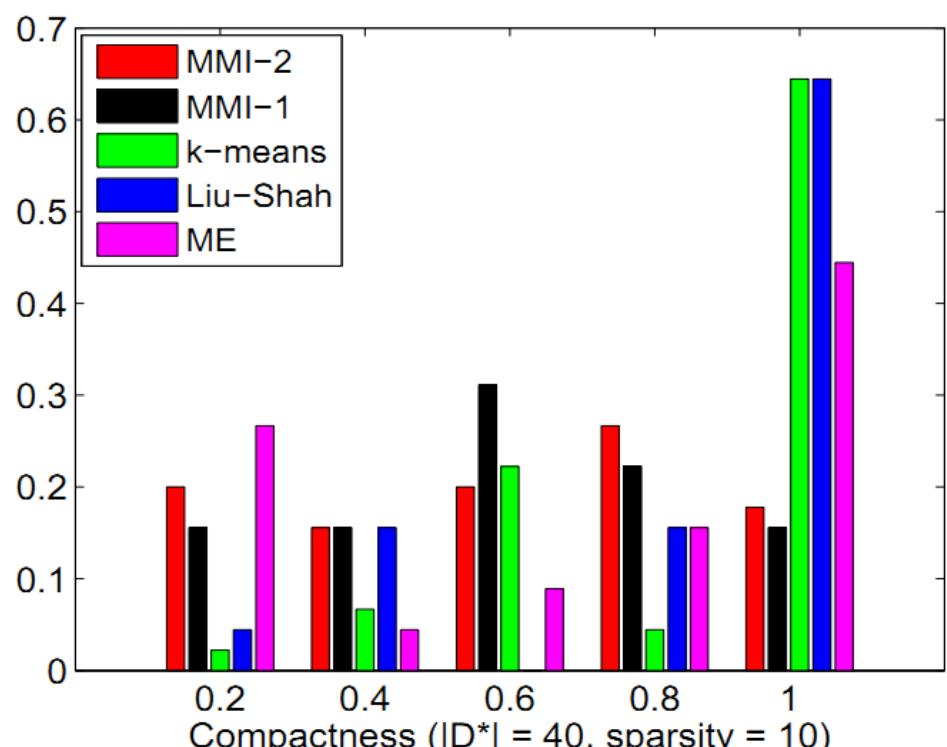
$$p(L|d^*) = \frac{p(d_1)}{p(d^*)} p(L|d_1) + \frac{p(d_2)}{p(d^*)} p(L|d_2)$$

$$p(d^*) = p(d_1) + p(d_2)$$

# Purity and Compactness

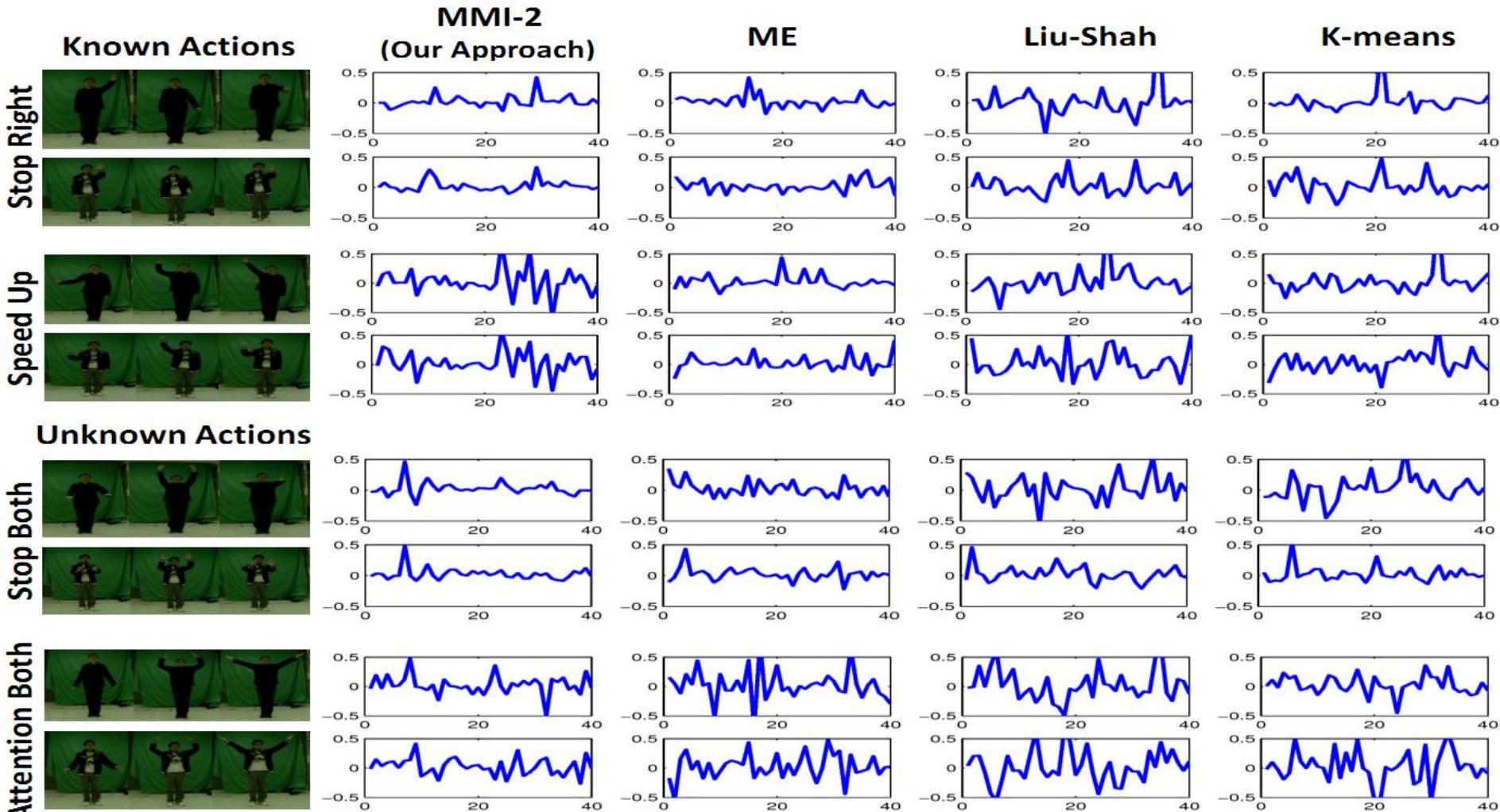


(a) Purity



(b) Compactness

# Representation Consistency



# Keck gesture dataset



Turn left

Turn right

Attention left

Attention right

Attention both

Stop left

Stop right



Stop both

Flap

Start

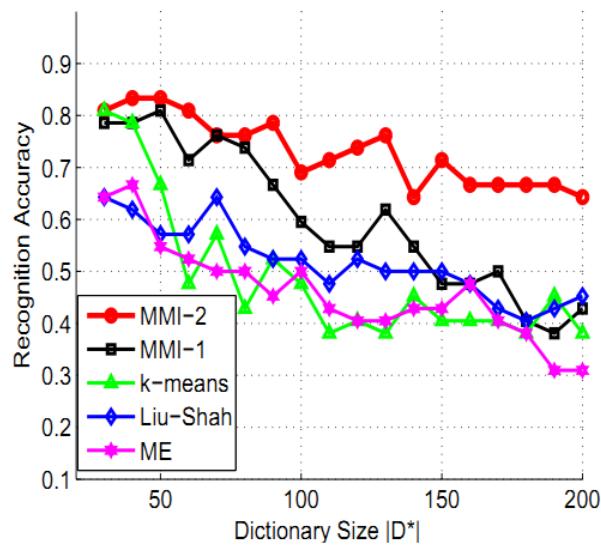
Go back

Close distance

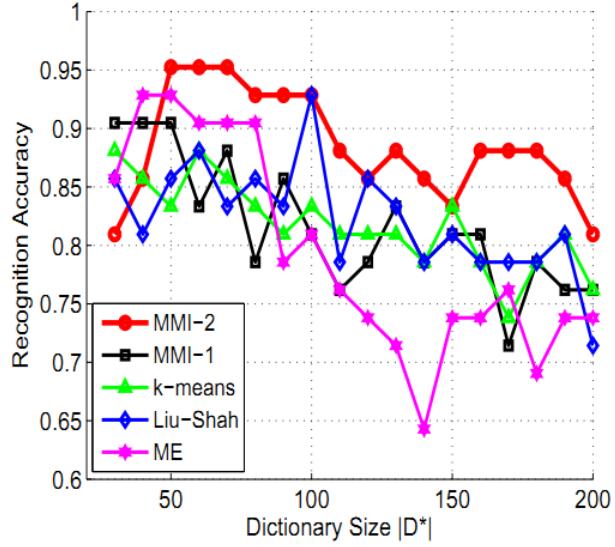
Speed up

Come near

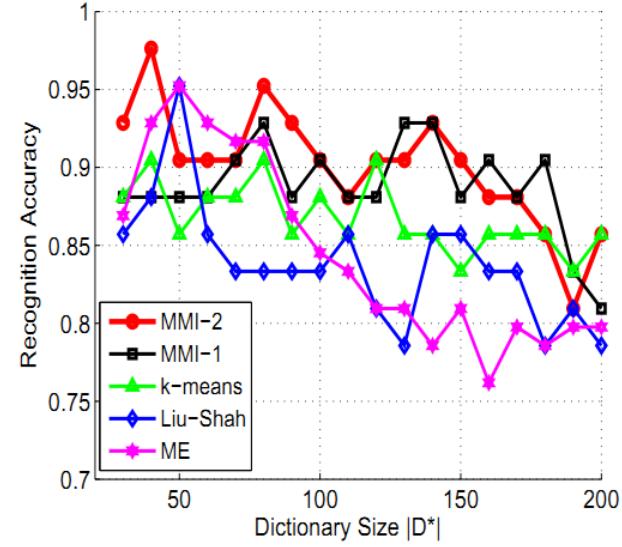
# Recognition Accuracy



(a) Shape ( $|D^o| = 600$ )



(b) Motion ( $|D^o| = 600$ )



(c) Shape and Motion ( $|D^o| = 1200$ )

The recognition accuracy using initial dictionary  $D^o$ : (a) 0.23 (b) 0.42 (c) 0.71