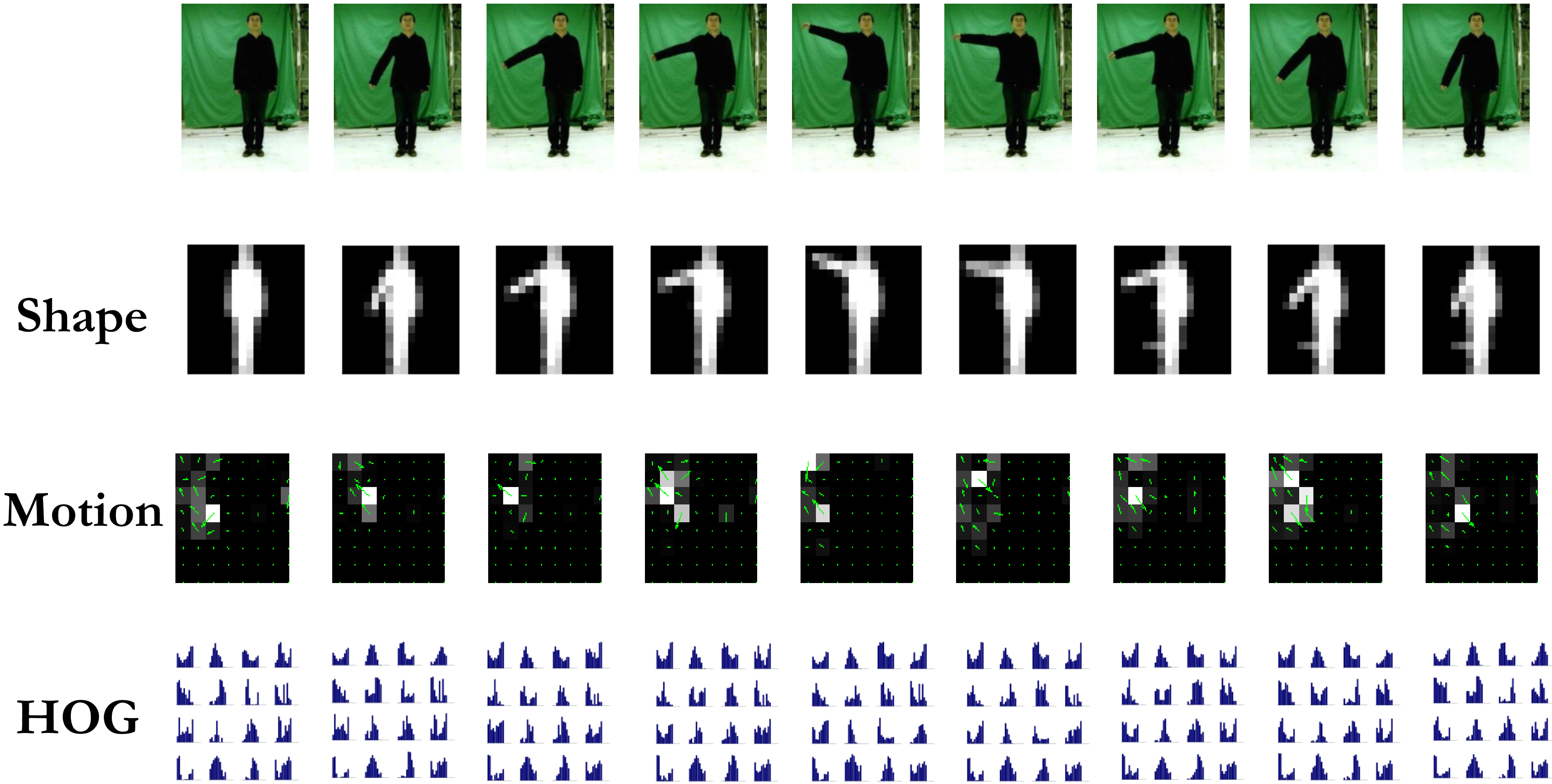
Sparse Dictionary-based Representation and Recognition of Action Attributes

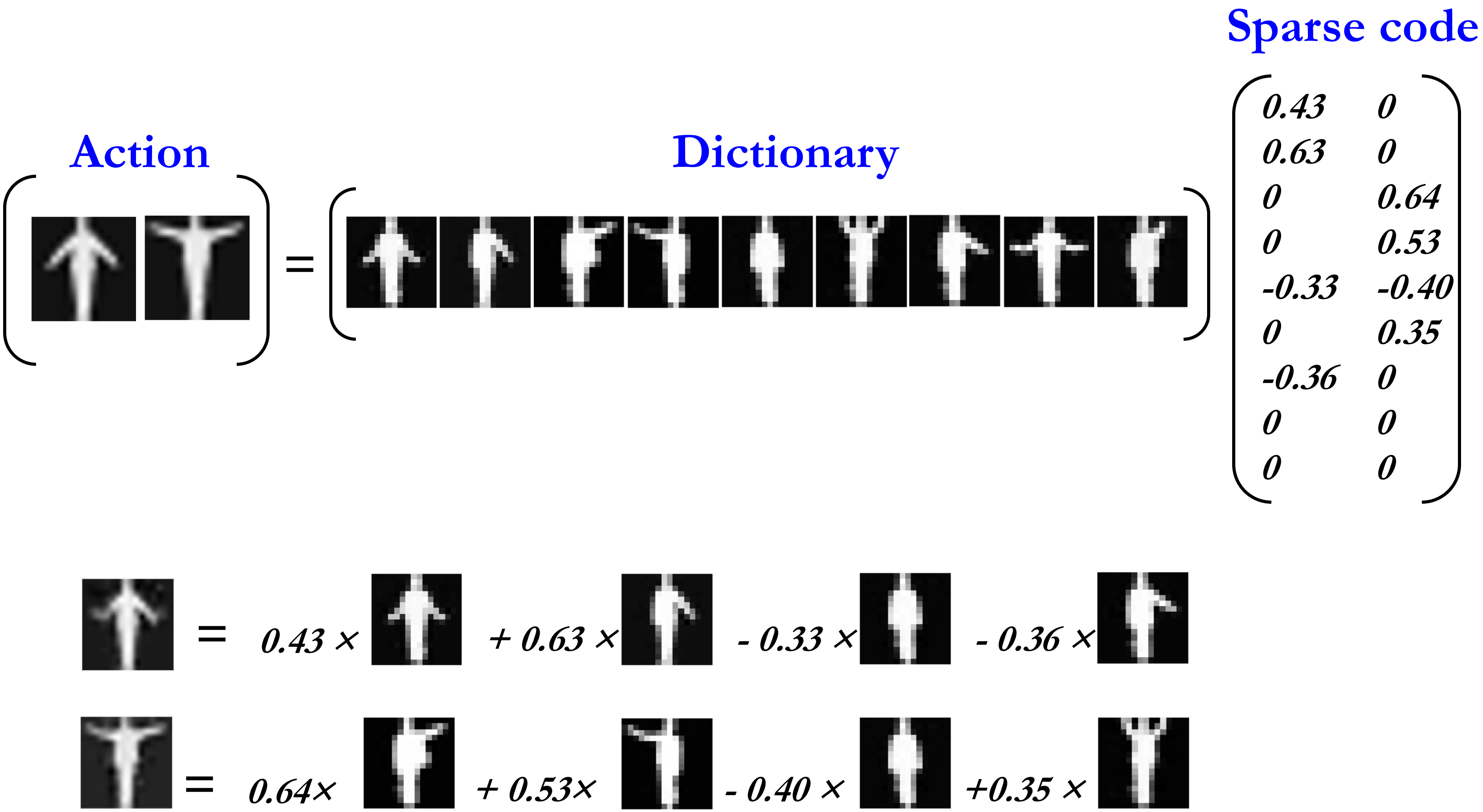
Qiang Qiu, Zhuolin Jiang, Rama Chellappa

Center for Automation Research,
Institute for Advanced Computer Studies
University of Maryland, College Park
qiu@cs.umd.edu, {zhuolin, rama}@umiacs.umd.edu

Action Feature Representation



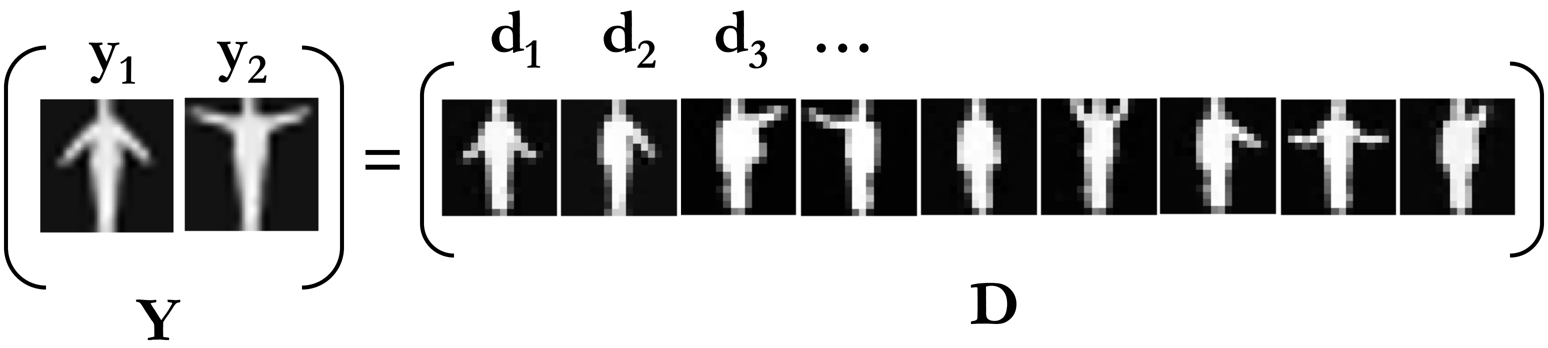
Action Sparse Representation



K-SVD

Input signals

Dictionary



Sparse codes

$$X = \begin{pmatrix} x_1 & x_2 \\ 0.43 & 0 \\ 0.63 & 0 \\ 0 & 0.64 \\ 0 & 0.53 \\ -0.33 & -0.40 \\ 0 & 0.35 \\ -0.36 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

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

- **K-SVD [1]**
  - **Input: signals Y, dictionary size, sparisty T**
  - **Output: dictionary D, sparse codes X**

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

Objective

Learn a

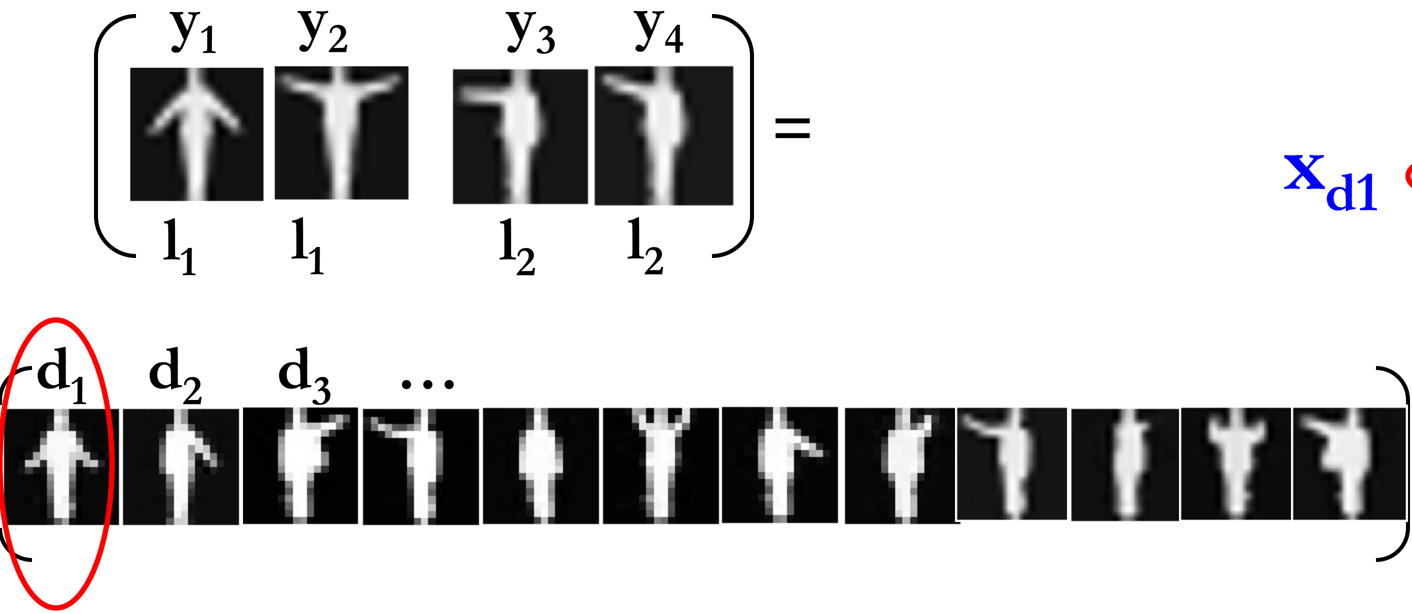
Compact and **Discriminative**

Dictionary.

Probabilistic Model for Sparse Representation

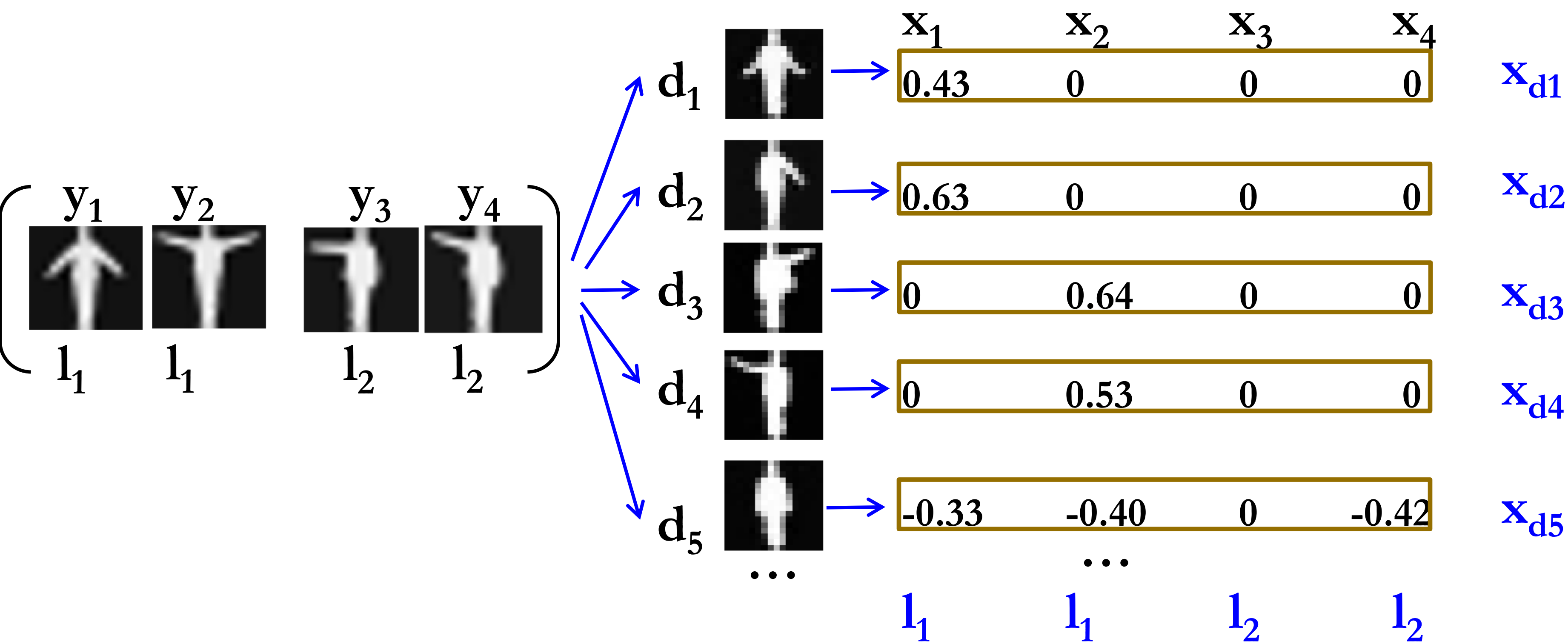
- A Gaussian Process
- Dictionary Class Distribution

More Views of Sparse Representation



	x_1	x_2	x_3	x_4
x_{d1}	*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*
	l_1	l_1	l_2	l_2

A Gaussian Process

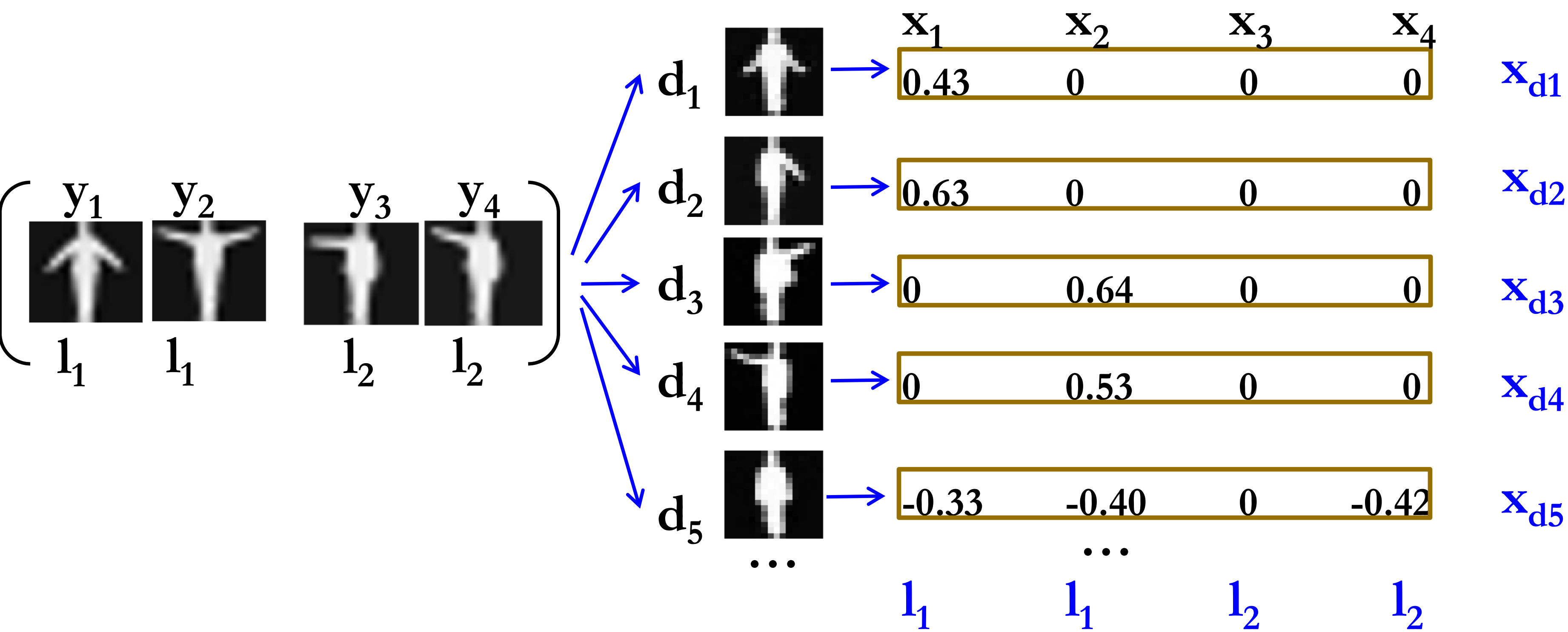


A Gaussian Process

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

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

Dictionary Class Distribution



Dictionary Class Distribution

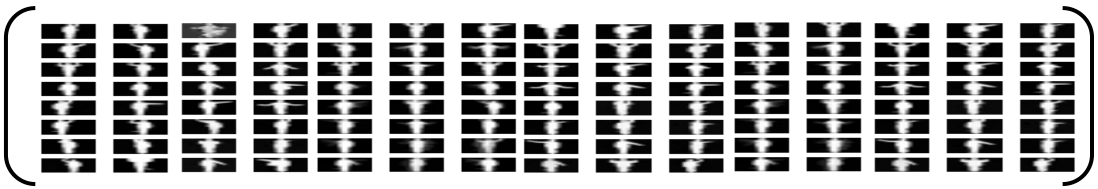
- $P(L|d_i), \ L \in [1, M]$
  - aggregate $|x_{di}|$ based on class labels to obtain a M sized vector
    - $P(L=l_1|d_5) = (0.33+0.40)/(0.33+0.40+0.42) = 0.6348$
    - $P(L=l_2|d_5) = (0+0.42)/(0.33+0.40+0.42) = 0.37$

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

$D^o =$ 

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

$$d^* = \arg\max_{d} H(d|D^*)$$

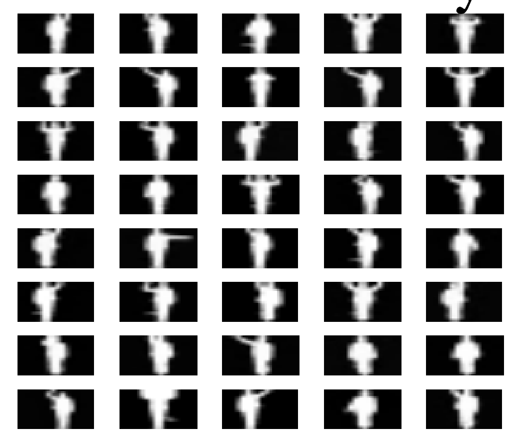
Where

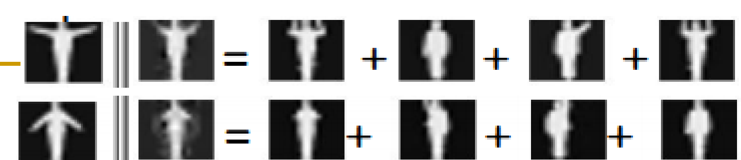
$$H(d^*|D^*) = \frac{1}{2} log(2\pi e \mathbb{V}(d^*|D^*))$$

- A good approximation to ME criteria

$$\arg\max_{D} H(D)$$

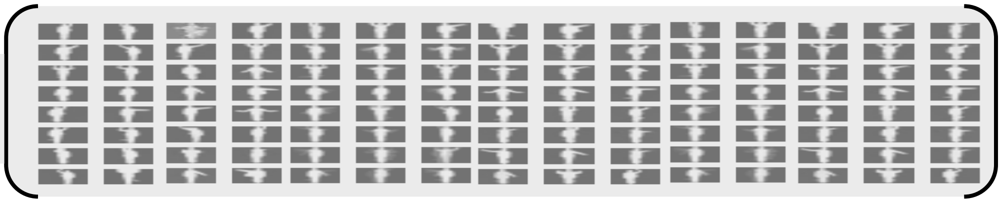
ME dictionary





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

- Initialize dictionary using k-SVD

$D^o =$ 

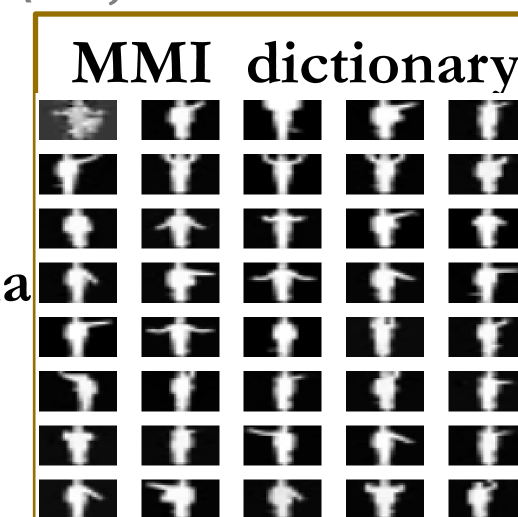
- Start with $D^* = \phi$
- Untill $|D^*| = k$, iteratively choose d^* from $D^o \backslash D^*$,

$$d^* = \arg\max_{d} H(d|D^*) - H(d|D^o \backslash (D^* \cup d))$$

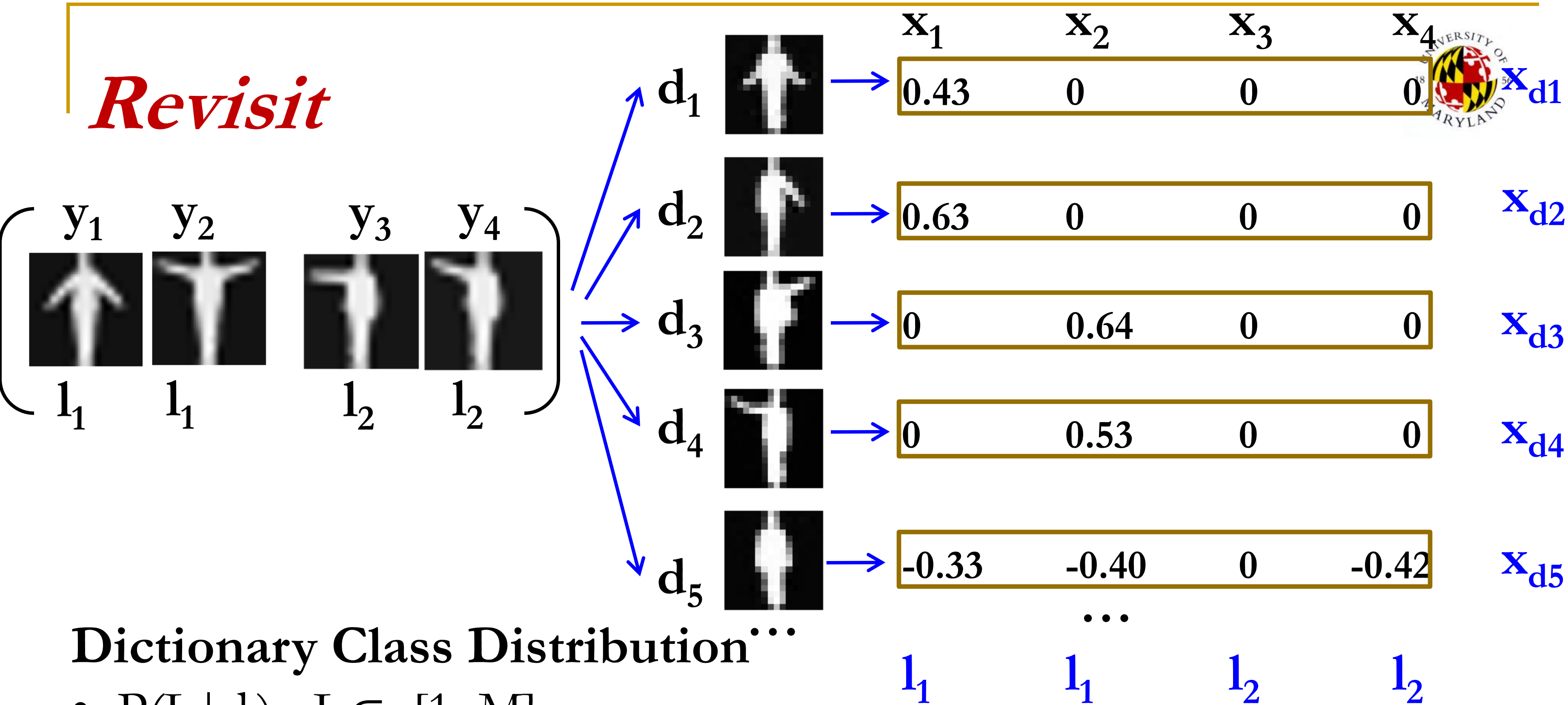
-A near-optimal approximation to MMI criteria

$$\arg\max_{D} I(D; D^o \backslash D)$$

Within (1-1/e) of the optimum



Revisit

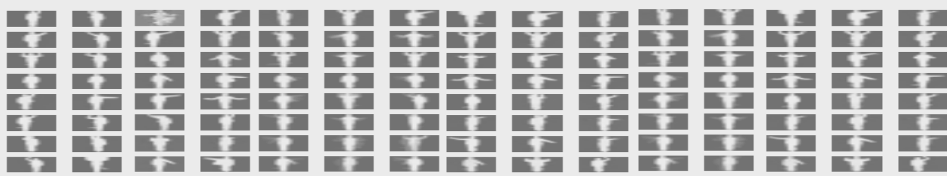


Dictionary Class Distribution

- $P(L|d_i)$, $L \in [1, M]$
  - 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

$D^o =$ 

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

$$d^* = \arg\max_{d} \left[H(d|D^*) - H(d|D^o \backslash (D^* \cup d)) \right] + \lambda \left[H(L_d|L_{D^*}) - H(L_d|L_{D^o \backslash (D^* \cup d)}) \right]$$

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

Other learning methods

- **K-means**
- **Liu-shah [1]**

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

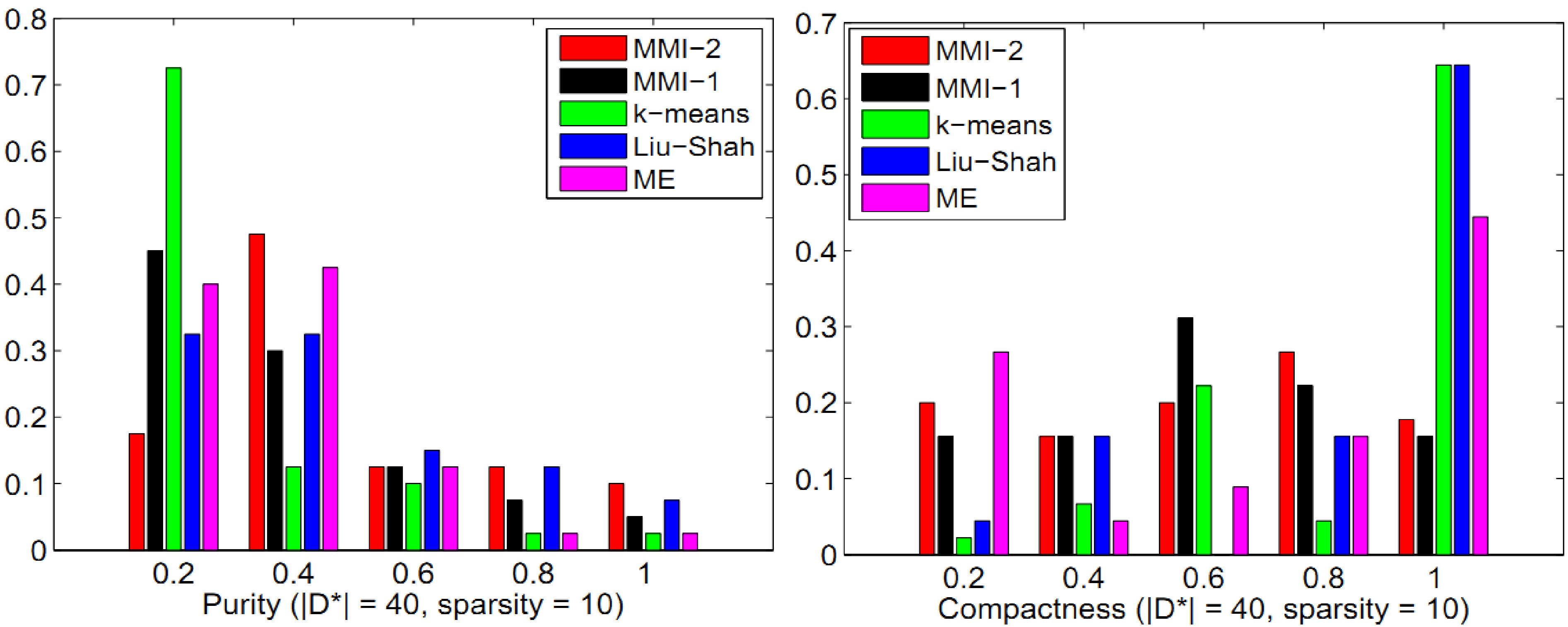
where

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

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

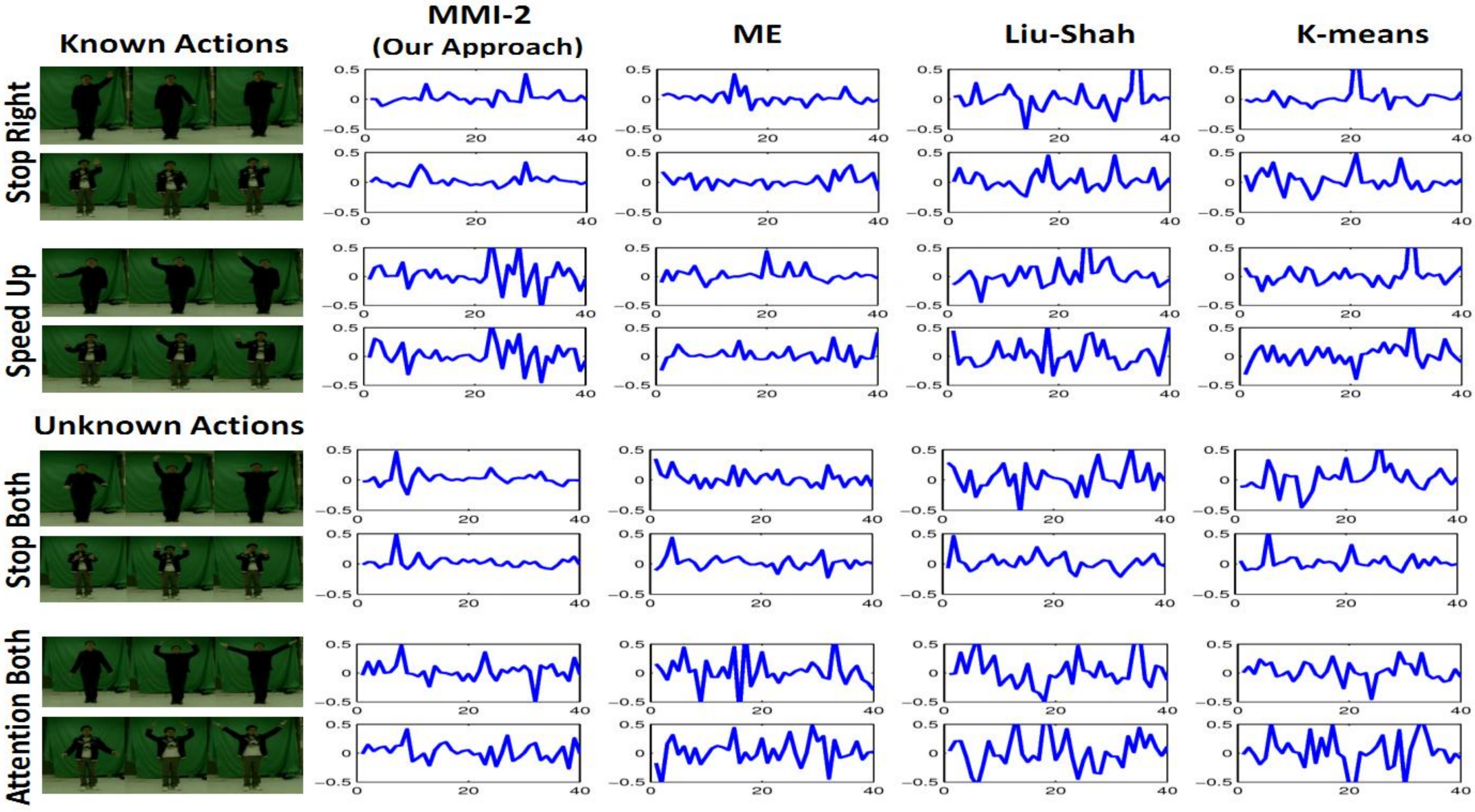
[1] J. Liu and M. Shah, Learning Human Actions via Information Maximization, CVPR 2008

Purity and Compactness

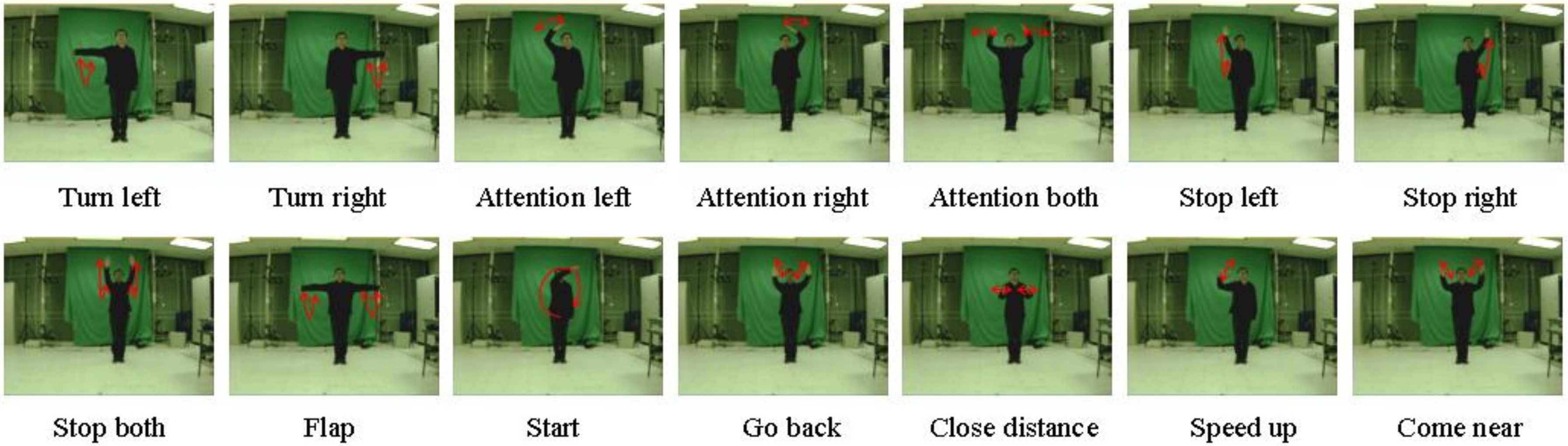


(a) Purity (b) Compactness

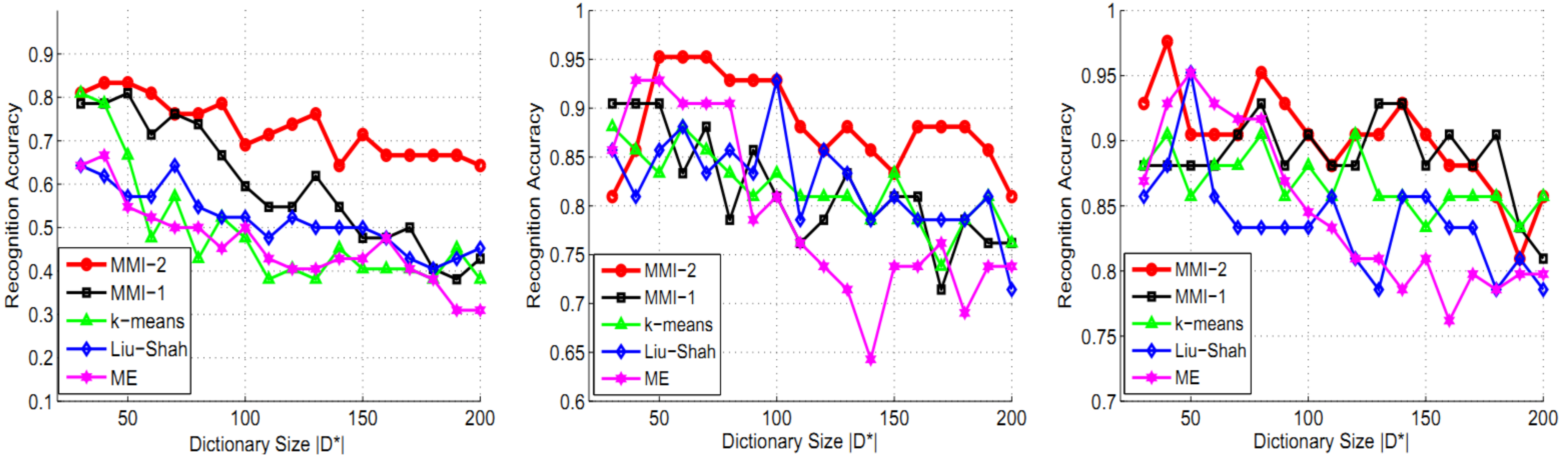
Representation Consistency



Keck gesture dataset



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