Self-Explanatory Sparse Representation for Image Classification

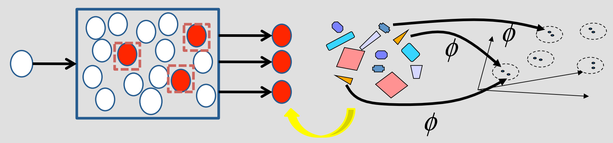
Bao-Di Liu*, Yu-Xiong Wang*, Bin Shen**, Yu-Jin Zhang, Martial Hebert

* contributed equally; ** the corresponding author

MOTIVATION

- Powerful tools in discovering data hidden structure

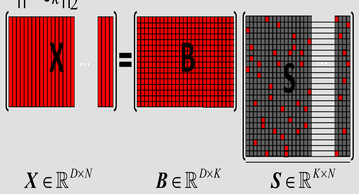
Sparse Representation | Kernel Methods (Multiple Kernel Learning)



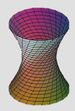
- Apply Kernel Trick to Sparse Representation?
  - Naïve solution:

$$\min_{B,S} f(B,S)=\|X-BS\|_F^2+2\alpha\|S\|_1 \quad s.t.\ \|B_{\bullet k}\|_2\le 1,\ \forall k=1,2,\cdots,K$$

$$\min_{B,S} f(B,S)=\|\phi(X)-\phi(B)S\|_F^2+2\alpha\|S\|_1 \quad s.t.\ \phi^T(B_{\bullet k})\phi(B_{\bullet k})\le 1,\ \forall k=1,2,\cdots,K$$



$X\in\mathbb{R}^{D\times N}$ $B\in\mathbb{R}^{D\times K}$ $S\in\mathbb{R}^{K\times N}$



Kernels:
Hellinger
HIK
Polynomial
RBF
…

- *However*, restricted…
  only specific kernels | approximation

SELF-EXPLANATORY REFORMULATION

- New Benefits I

$$\min_{B,S} f(B,S)=\|X-BS\|_F^2+2\alpha\|S\|_1 \quad s.t.\ \|B_{\bullet k}\|_2\le 1,\ \forall k=1,2,\cdots,K$$

- Conceptual interpretability
- More flexible dictionary design
- Easy for kernelization

$B=XW$

- New Benefits II

$$\min_{W,S} f(W,S)=\|X-XWS\|_F^2+2\alpha\|S\|_1 \quad s.t.\ \|XW_{\bullet k}\|_2\le 1,\ \forall k=1,2,\cdots,K$$

- Computational tractability
- Handle arbitrary kernels

$$\min_{W,S} f(W,S)=\|\phi(X)-\phi(X)WS\|_F^2+2\alpha\|S\|_1 \quad s.t.\ \|\phi(X)W_{\bullet k}\|_2\le 1,\ \forall k=1,2,\cdots,K$$

CONCLUSIONS

- Readily generalized into reproducing kernel Hilbert spaces for arbitrary kernels
- Capable of identifying various nonlinear structure information and sparse active components
- Enhanced visual representation ability, where SSSR-Hellinger & MSSR work better

SINGLE- & MULTIPLE- VIEW SSR

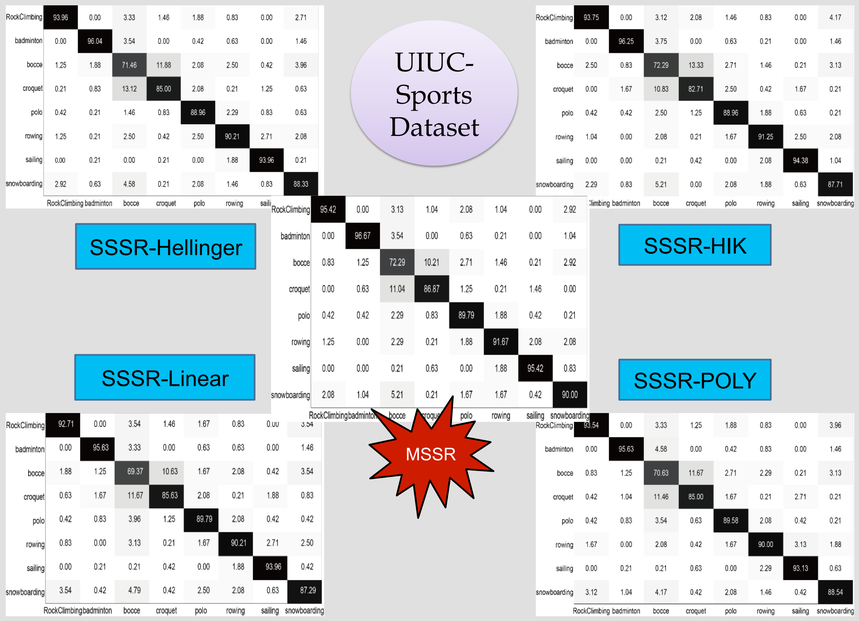
- SSSR

$$\min_{W,S} f(W,S)=\|\phi(X)-\phi(X)WS\|_H^2+2\alpha\|S\|_1 \quad s.t.\ \|\phi(X)W_{\bullet k}\|_H\le 1,\ \forall k=1,2,\cdots,K$$

- MSSR

$$\min_{W,S} f(W,S)=\sum_{H_g\in H}\|\phi^g(X)-\phi^g(X)W^gS^g\|_{H_g}^2+2\alpha\|S^g\|_1 \quad s.t.\ \|\phi^g(X)W^g_{\bullet k}\|_{H_g}\le 1,\ \forall k=1,2,\cdots,K;\ g=1,2,\cdots,G.$$

EXPERIMENTAL RESULTS



Caltech-256 Dataset

Methods	15 train	30 train	45 train	60 train
Single Scale SIFT				
ScSPM(1024) [Yang et al. 2009]	27.73±0.51	34.02±0.35	37.46±0.55	40.14±0.91
LLC(1024) [Wang et al. 2010]	27.74±0.32	32.07±0.24	35.09±0.44	37.79±0.42
KSR(1024) [Gao et al. 2010]	29.77±0.14	35.67±0.10	38.61±0.19	40.30±0.22
SCSR(1024) [Liu et al. 2013]	29.23±0.38	35.51±0.32	38.68±0.29	41.05±0.42
DLSM(1024) [Liu et al. 2013]	29.31±0.58	35.12±0.34	37.62±0.57	39.96±0.62
DLMM(1024) [Liu et al. 2013]	30.35±0.42	36.22±0.33	38.97±0.56	41.09±0.44
Ours (SSSR)				
Hellinger+1024+linearSVM	32.74±0.35	39.68±0.33	43.18±0.41	45.33±0.34
HIK+1024+linearSVM	32.38±0.47	39.13±0.48	42.40±0.34	44.86±0.32
POLY+1024+linearSVM	31.58±0.22	38.32±0.32	41.74±0.47	44.24±0.43
Linear+1024+linearSVM	31.52±0.31	38.19±0.33	41.39±0.49	43.95±0.63
Ours (MSSR)				
4096+linearSVM	34.06±0.36	41.14±0.43	44.72±0.42	47.26±0.43
4096+polySVM	**35.38±0.31**	**42.92±0.46**	**46.88±0.52**	**49.70±0.43**
Multiple Scale SIFT				
LLC(4096) [Wang et al. 2010]	34.36	41.19	45.31	47.68
KSRSPM-HIK(4096) [Gao et al. 2013]	33.61±0.34	40.63±0.22	44.41±0.12	47.03±0.35
IFK [Perronnin et al. 2010]	34.7±0.2	40.8±0.1	45.0±0.2	47.9±0.4
Ours (SSSR)				
Hellinger+4096+linearSVM	37.11±0.50	44.73±0.37	48.65±0.43	51.24±0.60
HIK+4096+linearSVM	35.95±0.36	43.45±0.29	47.27±0.33	49.96±0.56
POLY+4096+linearSVM	35.54±0.33	42.94±0.40	46.70±0.41	49.42±0.62
Linear+4096+linearSVM	35.66±0.43	43.10±0.28	46.98±0.38	49.52±0.60
Ours(MSSR)				
16384+linearSVM	37.12±0.41	44.95±0.38	48.89±0.37	51.47±0.72
16384+polySVM	**37.76±0.25**	**45.70±0.47**	**49.83±0.18**	**52.81±0.53**

OVERALL ALGORITHM

- Alternating Minimization

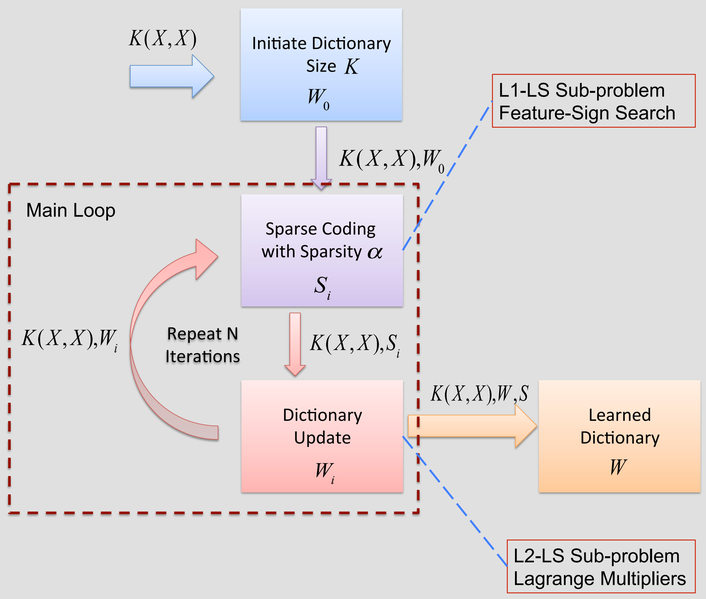
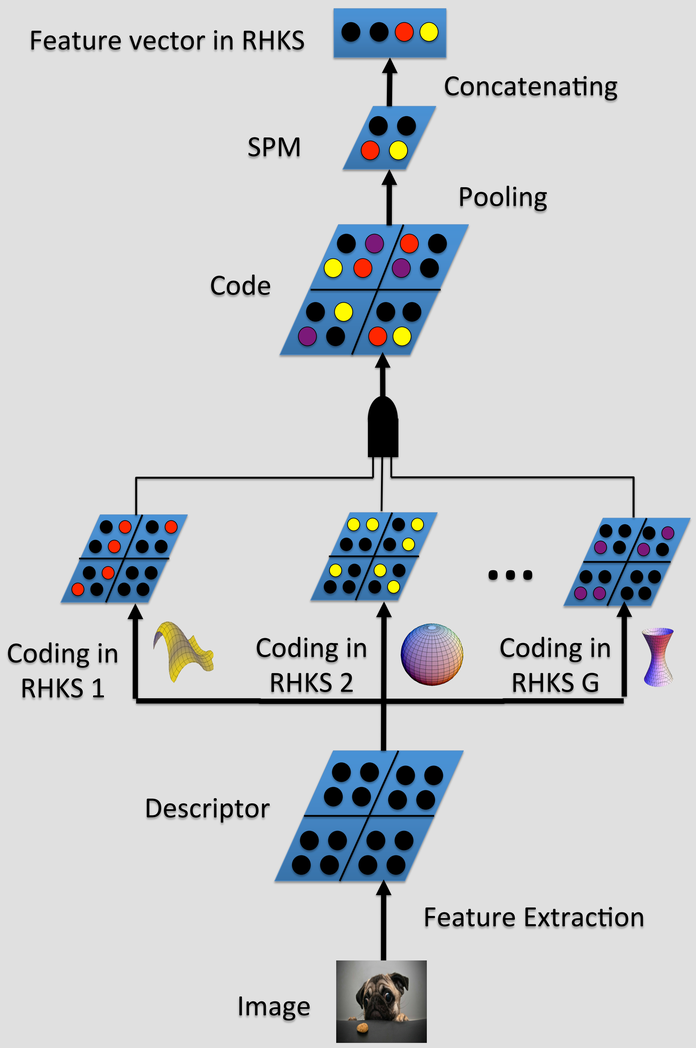



IMAGE CLASSIFICATION PIPELINE