Exploring Self-Similarity of Bag-of-Features for Image Classification

Chih-Fan Chen and Yu-Chiang Frank Wang — *Research Center for Information Technology Innovation, Academia Sinica, Taipei, Taiwan*

Typical Features of Use

BOF (Bag of Features) :

(1) Quantizes local image descriptors into distinct visual words.

(2) Uses a histogram representation to show the number of occurrences of each word for images.

SPM (Spatial Pyramid Matching) :

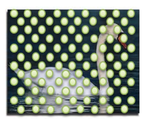
(1) Partitions an image into several grids in different scales.

(2) Concatenates the BOF form each grid to preserve the spatial infomation.

Overview of Image Classification

(1) Input Image

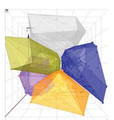
(2) Extracting Local Descriptors



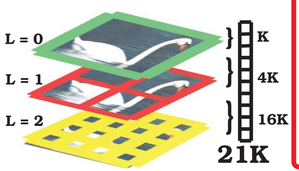
(3) Learning Codebook via Sparse Coding

$$\min_{\mathbf{A},\alpha} \sum_{i=1}^{N} \frac{1}{2}\|x_i - \mathbf{A}\alpha_i\|_2^2 + \lambda\|\alpha_i\|_1$$

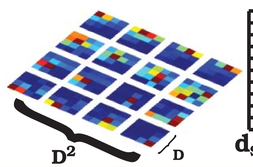
A is the codebook with K visual words



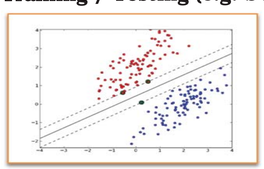
(4) Pyramid Max Pooling



(4') Learning of SSH

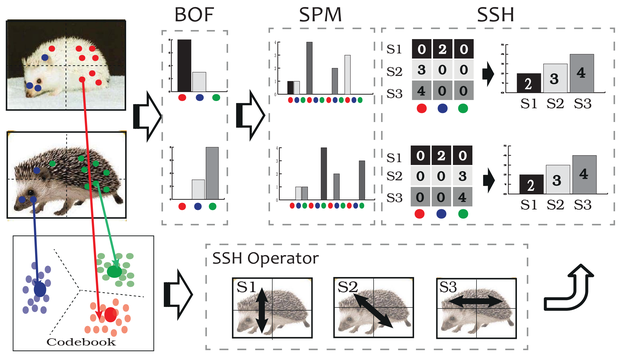


(5) Training / Testing (e.g. SVM)



Our Self-Similarity Hypercube (SSH) Model

We further exploit the structural information of an image by calculating the self-similarity of visual words.



Learning of SSH

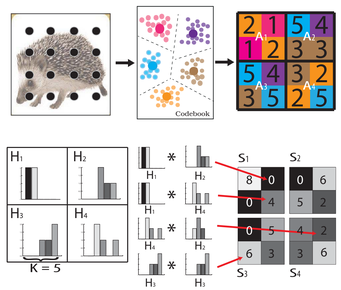
- **We calculate the inner products between each grid pair and form the SSH vector.**

  For example: (A1, A2) = 0 ; (A2, A3) = 2*1+1*1+1*2 = 5;

- **Original Size of SSH :** D^4.

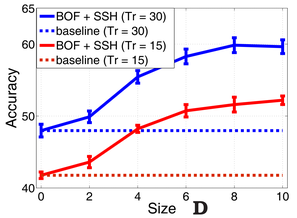
- **Since** $H_i^T H_j = H_j^T H_i$, $\mathbf{d}_{\mathbf{SSH}}$: $D^4 \rightarrow (D^2-1)D$

- **The SSH model :** $SSH = [S_{11}, S_{12}, ..., S_{1D^2}, ..., S_{(D^2-1)D}]$, *where* $S_{ij} = H_i^T H_j$

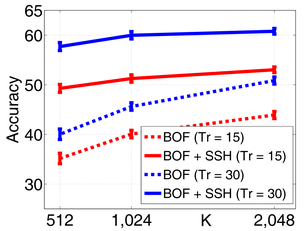


Experiments and Results

- **Performance comparisons on Caltech-101 with different SSH and codebook sizes.**
- **Note that D = 0 for BOF (baseline).**



- **We use D = 8 for our SSH, the dimension is 64 (64+1) = 2080.**



- **Our approach consistently achieved improve performance for different K.**

Type	BOF	BOF + SSH	SPM	SPM + SSH
Dimension	K (2048)	K+d_{SSH} (4128)	21K (43008)	21K + d_{SSH} (45088)
SIFT	51.5±1.8	64.4±0.8	73.3±0.6	74.7±0.9
HOG(8x8)	48.9±0.6	59.4±1.6	68.4±1.0	69.4±0.9
HOG(6x6)	39.5±1.6	54.2±1.3	64.3±0.9	66.3±0.6
HOG(4x4)	29.0±0.8	47.8±1.1	56.9±0.7	58.7±0.8

- **Our method consistently outperforms standard BOF and SPM.**
- **The dimensionality of our method is in the same order as BOF, while SPM is 21 times larger than BOF.**

Performance Comparisons

- **Comparisons of our SSH with state-of the-art methods on two image classification datasets.**

Database	Caltech 101		Caltech 256	
Training Images	15	30	15	30
Kernel Codebook (ECCV'08)	N / A	66.2±0.5	N / A	27.2±0.5
SPM (CVPR'06)	~59.0	67.6±1.4	~28.1	34.1±0.2
ScSPM (CVPR'09)	67.0±1.5	73.2±0.5	27.7±0.5	34.0±0.4
Ours	**68.2±0.7**	**74.7±0.9**	**33.0±0.3**	**39.7±0.3**

Selected object categories with improvement by SSH

- **Additional structural information of visual words provided by our SSH indeed exhibits complementary classification ability.**

Conclusions

- **We proposed SSH to exploit the self-similarity of visual words for image classification.**
- **SSH is constructed in the same feature domain as BOF.**
- **No eatra feature extraction / selection is needed.**
- **SSH provides additional and complementary representation and classification ability to BOF/SPM.**

Acknowledgement:

National Science Council,Taiwan (NSC 100-2221-E-001-018-MY2 and NSC 100-2631-H-001-013.)