Locality-constrained Group Sparse Representation for Robust Face Recognition

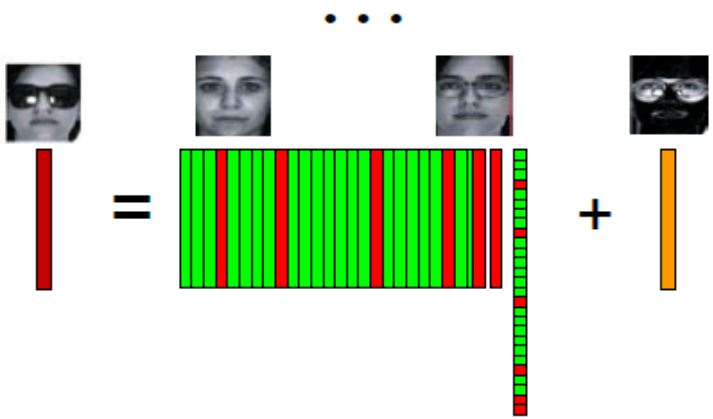
Yu-Wei Chao, Yi-Ren Yeh, Yu-Wen Chen, Yuh-Jye Lee and Yu-Chiang Frank Wang

Research Center for Information Technology Innovation
Academia Sinica, Taipei, Taiwan

Outline

Introduction

- Sparse Representation
  - ℓ_1-norm (Lasso)
  - $\ell_{1,2}$ mixed-norm (Group Lasso)



- Locality
  - Typically seen in problems of
    - Classification (e.g. NN, kNN)
    - Dimension reduction (e.g. Isomap, LLE)
    - Data representation (e.g. LLC [1])
  - Sparse representation does NOT guarantee locality [1]

[1] J.Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained Linear Coding for Image Classification. CVPR, 2010.

Our Proposed Method

- **Locality and Group-Sensitive Sparse Representation (LGSR)**
  - Learn data sparse representation by integrating both *group sparsity* and *data locality* into a unified formulation
  - Classification based on reconstruction error (e.g. MSE)
  - Possible extension to nonlinear versions by kernels

Different Coding Schemes

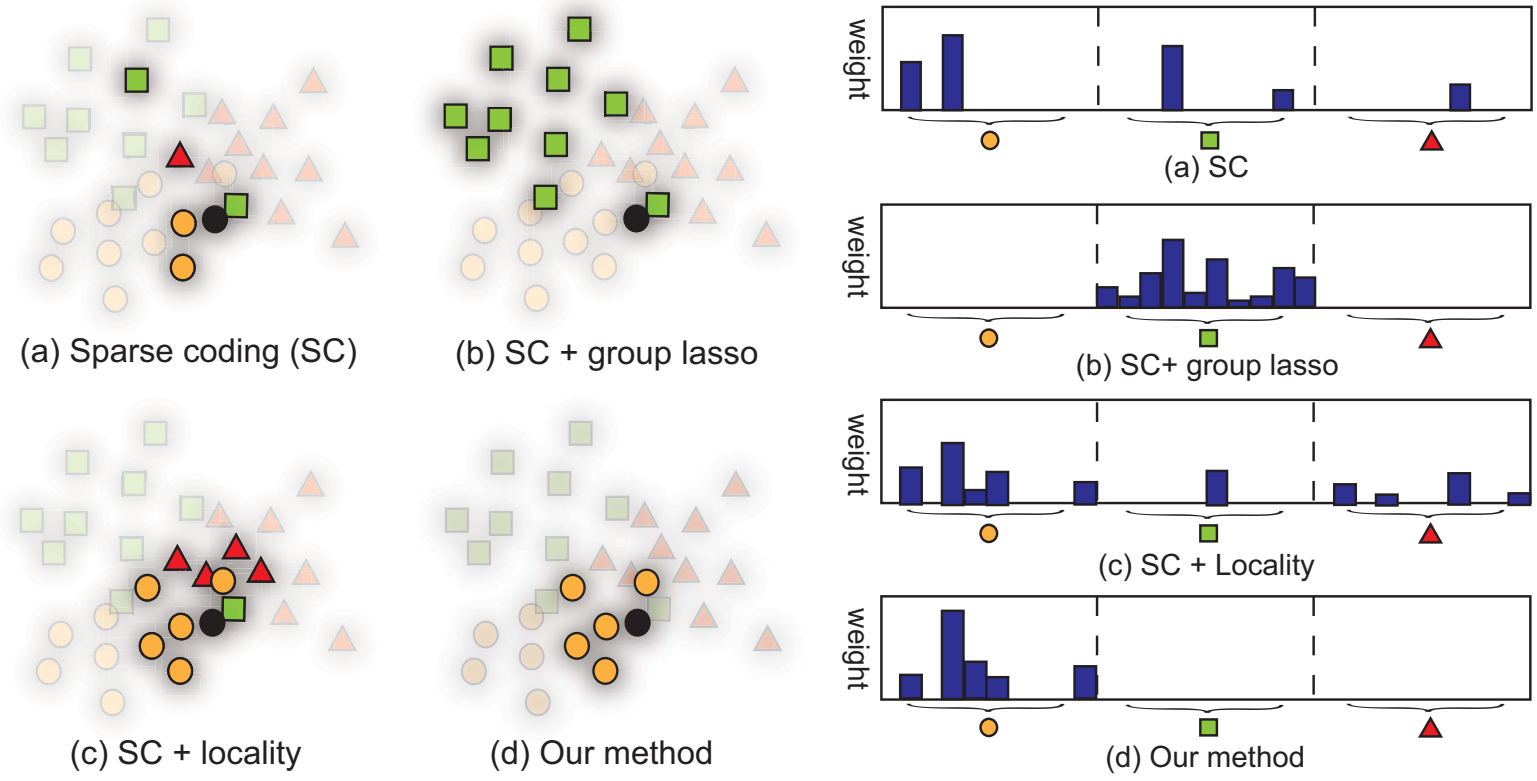


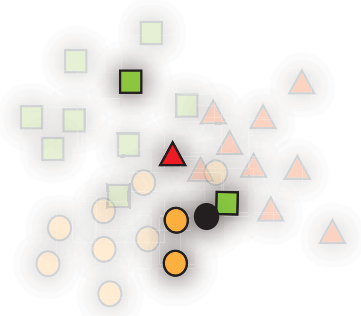
Figure: Data representation with different coding strategies

Outline

Image Sparse Representation for Classification

- Sparse coding (SC) aims to linearly reconstruct a data instance using an over-complete dictionary:

$$\min_{\mathbf{w}} \|\mathbf{x} - \mathbf{B}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1 . \quad (1)$$



Sparse coding (SC)

- $\mathbf{B}$ is the codebook (a set of visual words) and λ penalizes the ℓ_1-norm regularizer, which controls the sparsity of $\mathbf{w}$.
- For classification, one can [2]
  - Use each $\mathbf{w}$ as a new presentation of each $\mathbf{x}$
  - Train a discriminative classifier such as SVM

[2] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. CVPR, 2009.

Image Sparse Representation for Classification (cont'd)

- Another use of SC for Classification
  - Based on data reconstruction, Wright *et al.* proposed the sparse representation classification (SRC) method for face recognition. [3]
  - $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_c] \in \mathbb{R}^{m \times n}$: entire training set
    $\mathbf{A}_j = [\mathbf{x}_{j1}, \mathbf{x}_{j2}, \ldots, \mathbf{x}_{jn_j}] \in \mathbb{R}^{m \times n_j}$: training images from jth class
  - The SRC minimizes the image reconstruction error:

$$\min_{\mathbf{w}} \|\mathbf{x}_t - \mathbf{A}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1 , \tag{2}$$

    where $\mathbf{w} = [w_{11}, w_{12}, \ldots, w_{cn_c}]^\top \in \mathbb{R}^{n \times 1}$.
  - Classification by minimizing MSE:

$$j^* = \arg\min_j \|\mathbf{x}_t - \mathbf{A}_j \mathbf{w}_j\|_2^2 \tag{3}$$

  - Small MSE = Good Reconstruction = Good Recognition?

[3] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. IEEE PAMI, 2009.

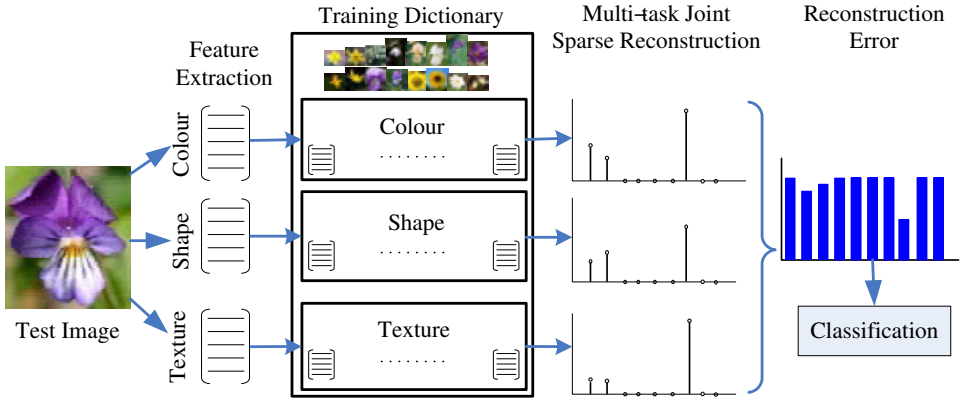
Outline

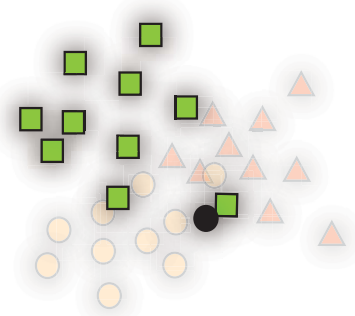
SC with the Group Lasso Constraint

- A joint sparsity representation: [4]

$$\min_{\mathbf{w}} \|\mathbf{x}_t - \mathbf{A}\mathbf{w}\|_2^2 + \lambda \sum_{j=1}^{c} \|\mathbf{w}_j\|_2. \tag{4}$$

where $\mathbf{w}_j = [w_{j1}, w_{j2}, \ldots, w_{jn_j}]^\top \in \mathbb{R}^{n_j \times 1}$ be the associated coefficients of $\mathbf{A}_j$.





SC + group lasso

- Data representation with group sparsity is achieved by posing a $\ell_{1,2}$ mixed-norm regularization.

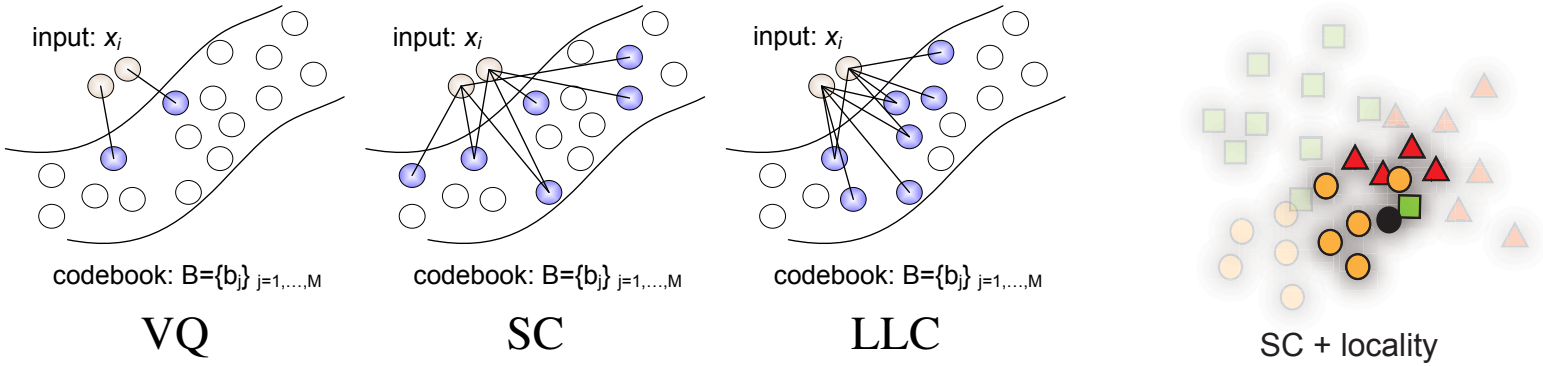
[4] X.-T. Yuan and S. Yan. Visual classification with multi-task joint sparse representation. CVPR, 2010.

SC with the Locality Constraint

- A locality-constrained linear coding (LLC) [1]
- LLC uses a distance regularization when minimizing MSE:

$$\min_{\mathbf{w}} \|\mathbf{x} - \mathbf{A}\mathbf{w}\|_2^2 + \lambda \|\mathbf{d} \odot \mathbf{w}\|_2^2 \,, \tag{5}$$

where $\mathbf{d} \in \mathbb{R}^{n \times 1}$ is the measurement of distance between $\mathbf{x}$ and each visual word in $\mathbf{A}$.

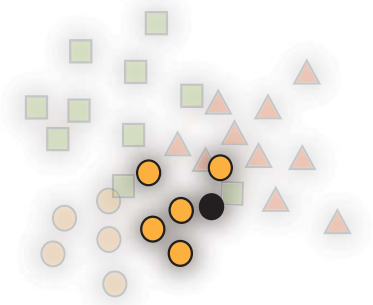


[1] J.Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained Linear Coding for Image Classification. CVPR, 2010.

Our LGSR Algorithm

- Integration of the $\ell_{1,2}$ mixed-norm regularization and locality constraint:

$$\min_{\mathbf{w}} \|\mathbf{x}_t - \mathbf{A}\mathbf{w}\|_2^2 + \lambda_1 \sum_{j=1}^{c} \|\mathbf{w}_j\|_2 + \lambda_2 \|\mathbf{d} \odot \mathbf{w}\|_2^2 ,$$



Our method

where λ_1 and λ_2 penalize the group sparsity and locality constraints, respectively.

- We adopt the distance metric in LLC:

$$d_{ji} = \exp\left(\frac{\|\mathbf{x}_t - \mathbf{x}_{ji}\|_2}{\sigma}\right). \tag{6}$$

- Classification by minimizing MSE:

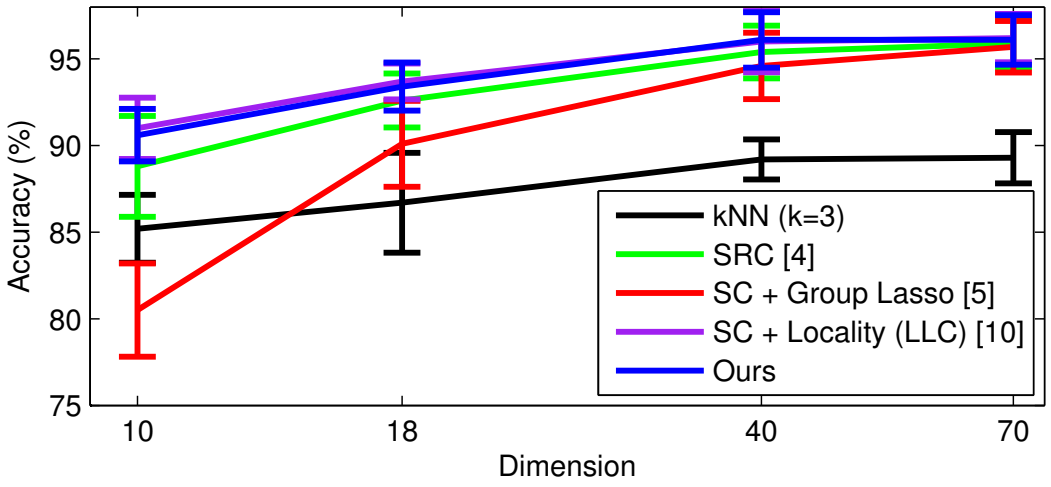
$$j^* = \arg\min_j \|\mathbf{x}_t - \mathbf{A}_j \mathbf{w}_j\|_2^2 \tag{7}$$

Outline

ORL Face Database

- 400 cropped face images of 40 human subjects under 10 different conditions (lighting, poses, etc.) with 56×46 pixels.
- For each subject, we randomly and equally split the data into training and test sets.
- We vary the number of Eigenfaces as features.
- For each experiment, we perform five random trials and compare our approach with SC-based methods.

Results of ORL Database (Eigenface)



dim	10	18	40	70
λ_1	0.1	0.05	0.01	0.1
λ_2	0.5	0.01	0.001	0.001

Figure: Top: Recognition using different SC approaches with different data dimensionality. Bottom: Parameters of our LGSR.

Results of ORL Database (Fisherface)

- We also extract Fisherfaces as features and repeat the experiments (all 39 eigenspaces for Fisherface are used).

Table: Recognition of the ORL database using Fisherface

kNN	SRC	SC+GL	SC+LLC	Ours
93.20±1.10	94.20±0.76	94.40±0.82	95.00±1.32	**95.40±1.43**

Extended Yale B Database (cropped)

- 2414 frontal images of 38 human subjects with gray-level 192 $\times$ 168 pixels, each image has up to 64 illumination variations.
- We extract Fisherfaces as the features (all 37 eigenspaces for Fisherface are used) and consider different number of training images (N_T) per class for evaluation.
- Once the training images are randomly extracted, the remaining will be test images.
- For each experiment, we perform three random trials.

Results of Extended Yale B Database

Table: Comparisons of recognition performance on Extended Yale B. N_T indicates the number of training images.

N_T	8	16	32
kNN (k=3)	76.70 ± 1.14	86.99 ± 1.36	93.29 ± 0.42
SRC	84.38 ± 1.21	92.95 ± 0.42	97.83 ± 0.14
SC+GL	84.39 ± 1.21	92.95 ± 0.42	97.83 ± 0.17
SC+LLC	84.74 ± 1.24	93.17 ± 0.19	97.89 ± 0.13
Ours	**85.17 ± 1.15**	**93.54 ± 0.40**	**98.14 ± 0.21**

Results of Extended Yale B Database (cont'd)

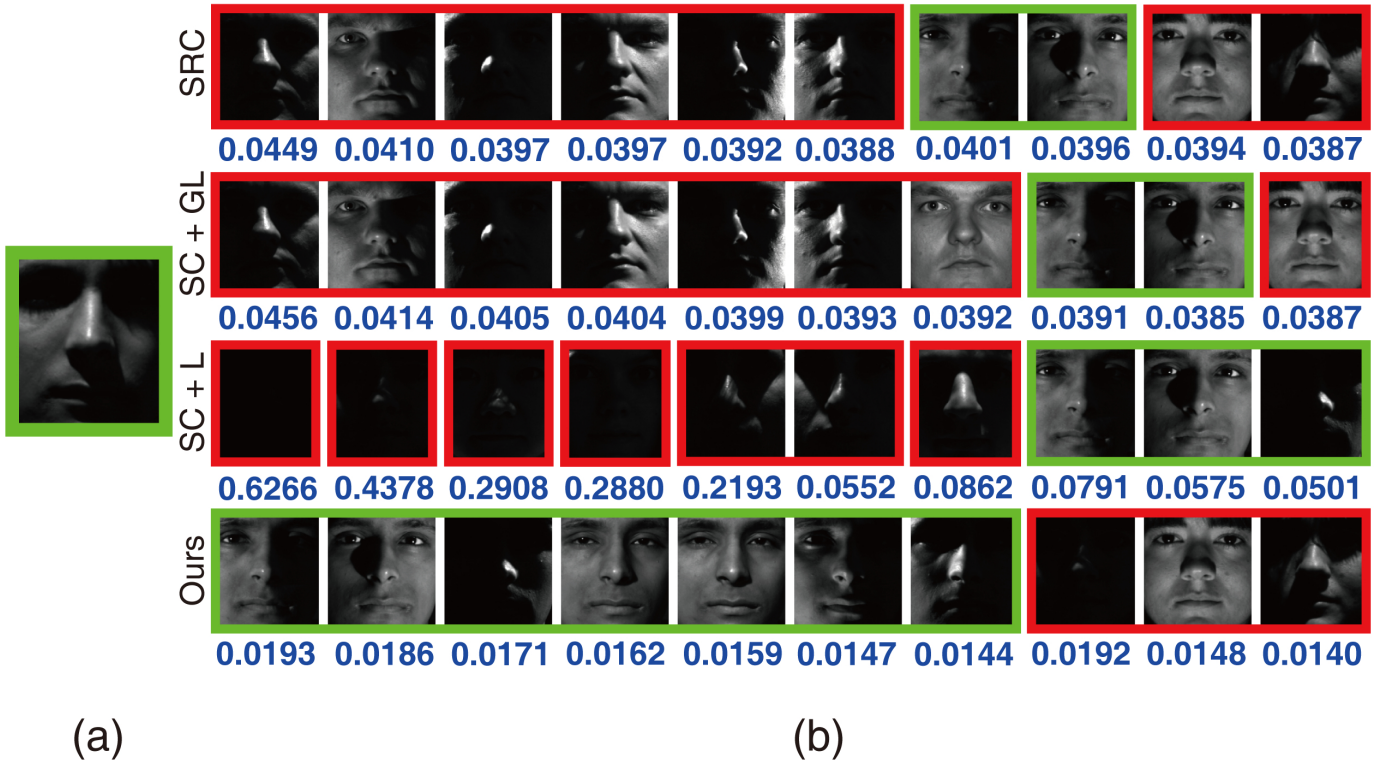


Figure: (a) An input test image. (b) Training images with non-zero weights using different SC methods (SC, SC+GL, SC+LLC, and our method).

Outline

Conclusion

- Our LGSR balances group sparsity and locality constraint and improves classification performance.
- More significant improvements were achieved when data with lower dimensionality was used.
- Our LGSR is training free.
- Future research directions at choosing **D** for encoding and optimization in large-scale problems.