A Novel and Robust Face Clustering Method via Adaptive Difference Dictionary

Jiaxiang Ren[1] , *Shengjie Zhao*[1] , *Kai Yang*[1] *and Brian Nlong Zhao*[2]

[1] Key Laboratory of Embedded System and Service Computing, Tongji University
[2] Shanghai High School International Division

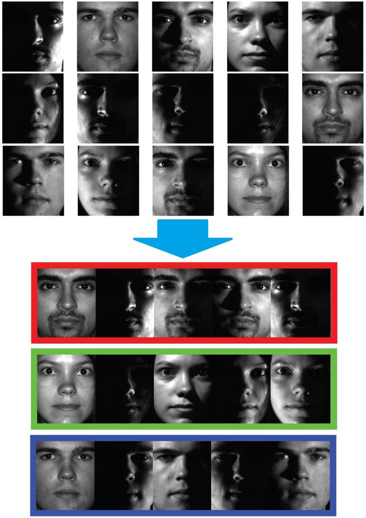
14 July, 2017

Contents

Sparse Subspace Clustering(SSC)

1. E. Elhamifar and R. Vidal, “Sparse subspace clustering”, *CVPR 2009. IEEE Conference on*, IEEE. pp. 2790–2797.
2. E. Elhamifar and R. Vidal, “Sparse subspace clustering: Algorithm, theory, and applications,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

Face Clustering



Input & Target

- **Input** : variant face images from multiple subjects
- **Target**:find images that belong to the same subject

The Extended Yale B Dataset

- images from 38 subjects
- 64 images per subject
- resolution: 192×168

SSC Algorithm

The Self-Expressiveness Property of the Data

Each data point in a union of subspaces can be efficiently reconstructed by a combination of other points in the dataset.

$$\begin{aligned} \min \quad & \|\boldsymbol{C}\|_1 + \lambda \|\boldsymbol{E}\|_1 \\ \text{s.t.} \quad & \boldsymbol{Y} = \boldsymbol{Y}\boldsymbol{C} + \boldsymbol{E}, \quad \operatorname{diag}(\boldsymbol{C}) = 0, \end{aligned} \tag{1}$$

- $\boldsymbol{C} = [\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N]$ is the **correlative coefficient matrix**
- $\boldsymbol{Y} = [\boldsymbol{Y}_{N_1}, \ldots, \boldsymbol{Y}_{N_K}] = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N] \in \mathbb{R}^{M \times N}$ is the input matrix, where $N = \sum_{k=1}^{K} N_k$
- $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N] \in \mathbb{R}^{M \times N}$ is the **auxiliary outliers matrix**

$$\boldsymbol{W} = |\boldsymbol{C}| + |\boldsymbol{C}^{\mathrm{T}}| \tag{2}$$

where $\boldsymbol{W} = |\boldsymbol{C}| + |\boldsymbol{C}|^{\mathrm{T}}$ is the **similarity matrix**, which means the similarity between the point i and j is equal to the sum of the absolute values of their correlative coefficients, *i.e.*, $|c_{ij}| + |c_{ji}|$.

Experiments Results

Algorithm	LSA	SCC	LRR	LRR-H	LRSC	SSC
2 Subjects						
Mean	32.80	16.62	9.52	2.54	5.32	**1.86**
Median	47.66	7.82	5.47	0.78	4.69	**0.00**
3 Subjects						
Mean	52.29	38.16	19.52	4.21	8.47	**3.10**
Median	50.00	39.06	14.58	2.60	7.81	**1.04**
5 Subjects						
Mean	58.02	58.90	34.16	6.90	12.24	**4.31**
Median	56.87	59.38	35.00	5.63	11.25	**2.50**
8 Subjects						
Mean	59.19	66.11	41.19	14.34	23.72	**5.85**
Median	58.59	64.65	43.75	10.06	28.03	**4.49**
10 Subjects						
Mean	60.42	73.02	38.85	22.92	30.36	**10.94**
Median	57.50	75.78	41.09	23.59	28.75	**5.63**

Figure: Clustering Error (%) of Different Algorithms on the Extended Yale B Dataset without Preprocessing the Data [1]

[1] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

Analysis of Results

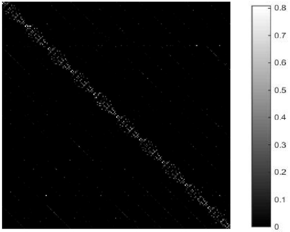


Figure: Coefficient matrix obtained when clustering error is less than 10%.

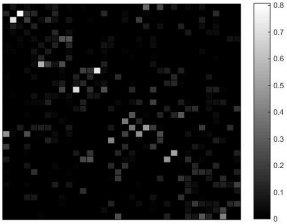


Figure: Coefficient matrix obtained when clustering error is **higher than 20%**.

The Defects of SSC

- Accuracy decreases for complicated variations
- Latent structures of multiple subspaces are too complicated to recover

Basic Idea of ESSC

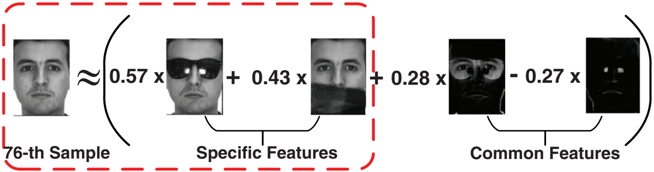


Figure: The sparse correlative coefficients of the 76-th sample recovered by the proposed ESSC.

Adaptive Difference Dictionary

- **Specific features** for clustering
- **Common features** for robustness
- More robust for complicated variations such as disguises (**improvement up to 9.0%**)
- Scalable and generalized for clustering **more subjects**

ESSC

Main Steps

1. Construction of the **adaptive difference dictionary**
2. Sparse optimization program
3. Spectral clustering

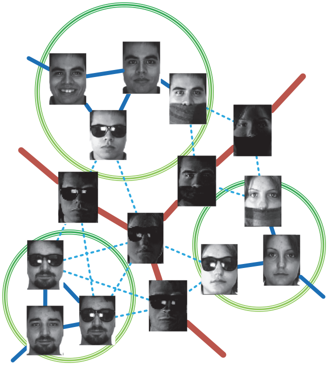


Figure: Face clustering with the adaptive difference dictionary. The adaptive differences play the role to **separate the samples** so that they can gather in their own subspaces.

Construction of the Adaptive Difference Dictionary

Computing coarse coefficient matrix:

$$\boldsymbol{Y} = \boldsymbol{YC} + \boldsymbol{E}, \quad \text{s.t.} \quad \text{diag}(\boldsymbol{C}) = \mathbf{0}. \tag{3}$$

Constructing the difference dictionary items:

$$\text{SCR}(\boldsymbol{c}_i) \triangleq \frac{\max(\boldsymbol{c}_i)}{\|\boldsymbol{c}_i\|}. \tag{4}$$

$$\begin{aligned} \boldsymbol{D} &\triangleq \{\boldsymbol{d}_* | \forall \text{SCR}(\boldsymbol{c}_*) > 0.1\} \in \mathbb{R}^{M \times N_d}, \\ \boldsymbol{d}_* &\triangleq \boldsymbol{y}_* - \boldsymbol{y}_{\max(\boldsymbol{c}_*)}, \end{aligned} \tag{5}$$

Sparse optimization program via the adaptive difference dictionary

Computing robust coefficient matrix:

$$\boldsymbol{Y} = [\boldsymbol{Y}\boldsymbol{D}] \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{B} \end{bmatrix} + \boldsymbol{Z}, \quad \text{s.t.} \quad \text{diag}(\boldsymbol{C}) = \mathbf{0}, \tag{6}$$

where $\boldsymbol{Z}$ models the Gaussian-noise in data. The corresponding constrained optimization program is

$$\begin{aligned} &\min \quad \left\| \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{B} \end{bmatrix} \right\|_1 + \frac{\lambda_z}{2} \|\boldsymbol{Z}\|_F^2 \\ &\text{s.t. } \boldsymbol{Y}{=}[\boldsymbol{Y}\boldsymbol{D}] \begin{bmatrix} \boldsymbol{C} \\ \boldsymbol{B} \end{bmatrix} {+}\boldsymbol{Z},\ \boldsymbol{C}^\top \mathbf{1}{=}\mathbf{1},\ \text{diag}(\boldsymbol{C}){=}\mathbf{0}, \end{aligned} \tag{7}$$

which can be solved using the **ADMM** approach. Thereafter, we use a spectral clustering to get the final clustering results.

Geometric Interpretation

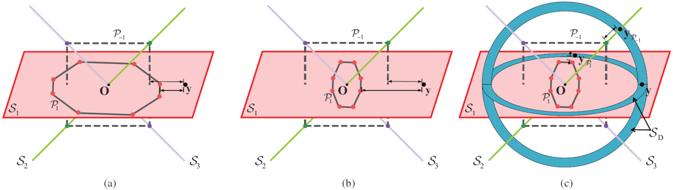


Figure: The sparse representation for recovering an image sample $\boldsymbol{y} \in \mathcal{S}_1$ in the intersection of $\mathcal{S}_1$ and $\mathcal{S}_2 \oplus \mathcal{S}_3$. (a) The distance to $\mathcal{P}_1$ is shorter than to $\mathcal{P}_{-1}$, so the sparse representation recovers correctly. (b) The distribution of the samples in $\mathcal{S}_1$ is odd because the spanned subspace is close to a line. The distance to $\mathcal{P}_1$ is larger than to $\mathcal{P}_{-1}$, so the sparse representation recovers incorrectly. (c) The adaptive difference dictionary generates **the common feature space** $\mathcal{S}_\mathrm{D}$, where any image sample can "**travel around**" to find the nearest polytope of the subspace correctly.

Clustering Variant Face Images

Table: Clustering Error Rates (%) of Different Algorithms on the AR Database Using Different Features for $K = 100$ Subjects

Variation Sample × Subject	Feature *(Dimension)*	Method			
		LRR	SSC	RPCA+SSC	ESSC
Expression 4×100	Downsample(55×40)	73.00	14.50	16.00	**13.00**
	LBP(5192)	70.75	8.75	**4.25**	10.00
Illumination 3×100	Downsample(55×40)	65.67	31.00	**30.33**	31.00
	LBP(5192)	67.67	6.00	6.00	**0.33**
Disguise 3×100	Downsample(55×40)	68.00	57.33	60.33	**55.00**
	LBP(5192)	65.33	17.67	14.33	**12.67**

The clustering error for ESSC is the **lowest** in almost all cases which confirms the effectiveness of the adaptive difference dictionary.

Clustering Scalability

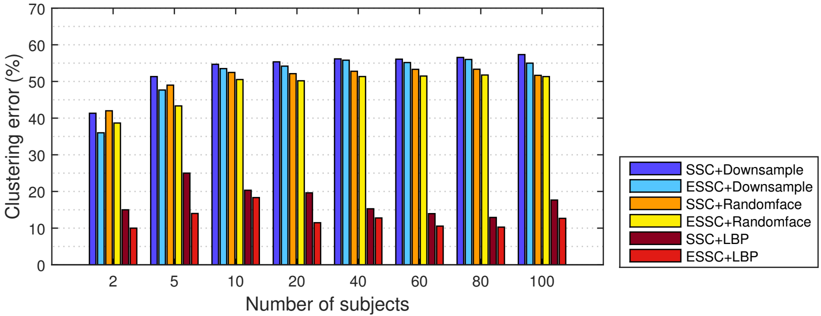


Figure: Clustering error rates for variant disguises on the AR database as a function of the number of subjects.

Q & A

Thanks!