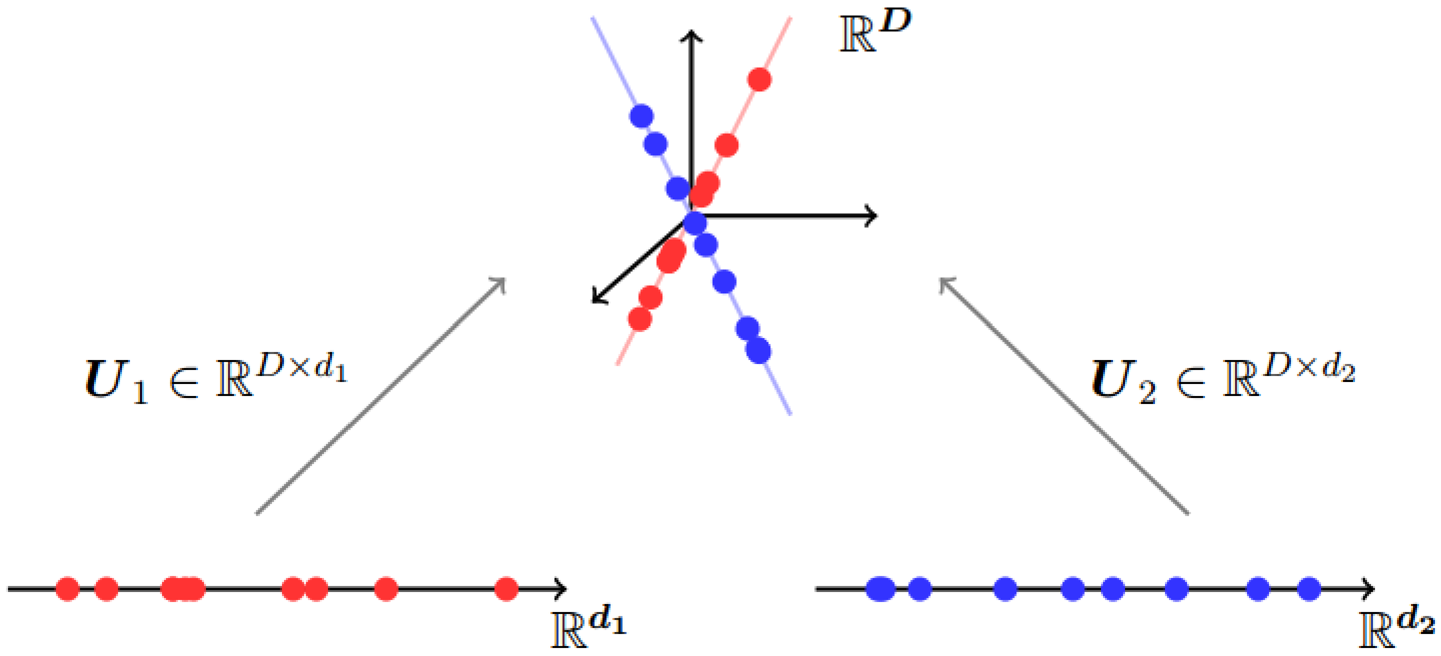
Sparse Subspace Clustering

Paper Presentation (EE698M)

Abhay Kumar

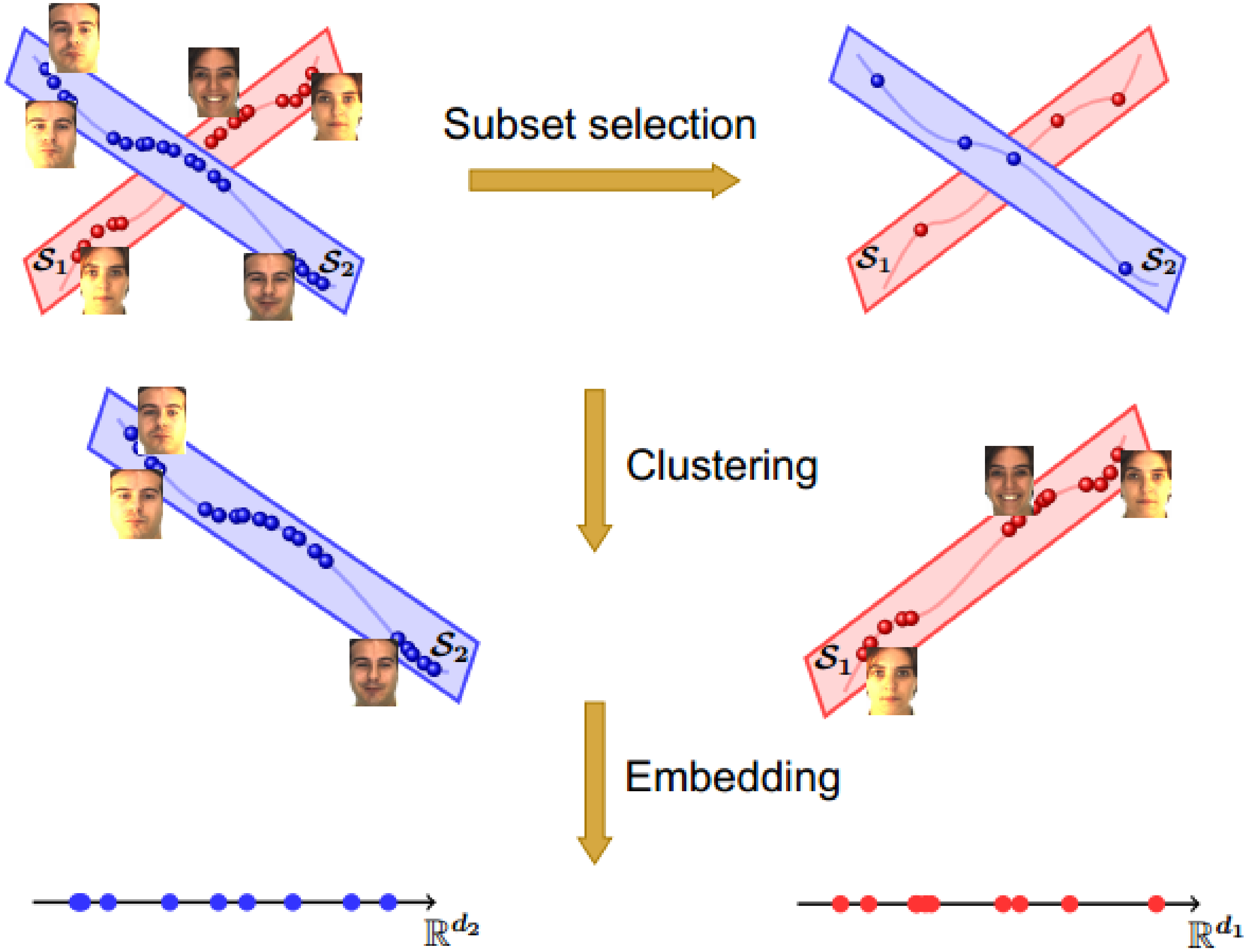
Subspace clustering

- Cluster data drawn from multiple low-dimensional linear or affine subspaces embedded in a high-dimensional space



Subspace clustering : Purpose

- Separate data into subspaces
- Find low-dimensional representations



Various Methodology:

- **K-subspaces**
  - Assigns points to subspaces →Fit subspace to each cluster →Iterate
  - Drawback: Requires Number and dimensions of subspaces to be known
- Statistical approaches such as **Mixture of Probabilistic PCA**, **Multi-stage Learning**
  - Assuming each subspace has Gaussian distribution →subspace estimation by EM
  - Drawback: Requires Number and dimensions of subspaces to be known
- Factorisation based methods
  - low-rank factorization of the data matrix
  - segmentation by thresholding the entries of a similarity matrix
- Generalized Principal Component Analysis (**GPCA**)
  - fit the data with a polynomial whose gradient at a point gives a vector normal to the subspace containing that point
- Information theoretic approaches, such as Agglomerative Lossy Compression (**ALC**)
  - Model each subspace as degenerate Gaussian→ segment data so as to minimise the coding length needed to fit these points with the mixture of Gaussians

Challenges:

- Intersecting subspaces
- noise, outliers, missing entries
- Computational complexity: NP hard (non-deterministic polynomial-time)
- <u>Knowledge of dimension/number of subspaces</u>

Sparse representation in a single subspace

- Sparse representation in a single subspace

$$x = \sum_{i=1}^{D} s_i \psi_i = \Psi s$$ where x in $\mathbb{R}^D$, $\Psi = [\psi_1, \psi_2, \cdots, \psi_D]$ $\{\psi_i \in \mathbb{R}^D\}_{i=1}^{D}$.

- In many cases x can have a sparse representation in a properly chosen basis Ψ.
- we do not measure x directly. Instead, we measure m linear combinations of entries of x of the form $y_i = \phi_i^\top x$ $i \in \{1, 2, \cdots, m\}$
- $y = [y_1, y_2, \cdots, y_m]^\top = \Phi x = \Phi \Psi s = A s$

  where $\Phi = [\phi_1, \phi_2, \cdots, \phi_m]^\top \in \mathbb{R}^{m \times D}$ is called the measurement matrix.
- one can recover K-sparse signals/vectors if $K \lesssim m / \log(D/m)$.
- Optimisation problem: $\min \|s\|_0$ subject to $y = As$.

Sparse representation in a union of subspaces

- Let $\{A_i \in \mathbb{R}^{D\times d_i}\}_{i=1}^{n}$: set of bases for n disjoint linear subspaces
- $\boldsymbol{y} = A\boldsymbol{s} = [A_1,\ A_2, \cdots,\ A_n]\,[\boldsymbol{s}_1^\top, \boldsymbol{s}_2^\top, \cdots, \boldsymbol{s}_n^\top]^\top$
- What if y belong to i-th subspace ??
- Optimisation Problems:-

$$\min \sum_{i=1}^{n} 1(\|\boldsymbol{s}_i\|_2 > 0) \quad \text{subject to} \quad \boldsymbol{y} = A\boldsymbol{s}.$$

$$\min \sum_{i=1}^{n} \|\boldsymbol{s}_i\|_2 \quad \text{subject to} \quad \boldsymbol{y} = A\boldsymbol{s}$$

Clustering linear subspaces:

- Known:
  - Sparsifying basis for the union of subspaces given by the data matrix
- Unknown:
  - not have any basis for any of the subspaces
  - don’t know which data belong to which subspace
  - don’t know total number of subspaces

Subspace clustering

- Assume:
  - $\{\mathcal{S}_i\}_{i=1}^n$: n-independent linear subspaces (unknown ☹)
  - $\{\boldsymbol{y}_j \in \mathbb{R}^D\}_{j=1}^N$: N data points collected from union of subspaces (known ☺)
  - $\{d_i \ll D\}_{i=1}^n$: unknown dimensions of n-subspaces.
  - $\{A_i \in \mathbb{R}^{D\times d_i}\}_{i=1}^n$: unknown bases for n-subspaces.
- Represent data matrix as $Y = [\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N] = [Y_1, Y_2, \cdots, Y_n]\Gamma \in \mathbb{R}^{D\times N}$

  where $Y_i \in \mathbb{R}^{D\times N_i}$; $N = \sum_{i=1}^n N_i$ and $\Gamma \in \mathbb{R}^{N\times N}$ is unknown permutation matrix that specifies the segmentation of data

Subspace clustering

- Let $s = \Gamma^{-1}[s_1^\top, s_2^\top, \cdots, s_n^\top]^\top$ where $s_i \in \mathbb{R}^{N_i}$
- If a point y is a new data point in $\mathcal{S}_i$?? → $s_i \neq 0$ and $s_j = 0$ for all $j \neq i$.
- Optimisation problem:

$$\min \|s\|_0 \quad \text{subject to} \quad y = Y s$$

Subspace clustering

- Let $Y_{\hat{i}} \in \mathbb{R}^{D \times N-1}$ be the matrix obtained from Y by removing i-th column, $\boldsymbol{y}_i$.

$$\min \|\boldsymbol{c}_i\|_1 \quad \text{subject to} \quad \boldsymbol{y}_i = Y_{\hat{i}} \boldsymbol{c}_i.$$

- The optimal solution $\boldsymbol{c}_i \in \mathbb{R}^{N-1}$ has non-zero entries corresponding to the columns in $Y_{\hat{i}}$ that lie in the same subspace as $\boldsymbol{y}_i$
- Insert zero at i-th row of $\boldsymbol{c}_i$ to make it N-dimensional $\hat{\boldsymbol{c}}_i \in \mathbb{R}^N$
- Solve for each point $i = 1, \ldots, N$
- Finally obtained a matrix of coefficients $C = [\hat{\boldsymbol{c}}_1, \hat{\boldsymbol{c}}_2, \cdots, \hat{\boldsymbol{c}}_N] \in \mathbb{R}^{N \times N}$

Subspace clustering

sparse representation comes from same subspace

$$\min \|\boldsymbol{c}_i\|_0 \quad \text{s.t.} \quad \boldsymbol{y}_i = \boldsymbol{Y}\boldsymbol{c}_i, \ c_{ii} = 0$$

⬇

$$\min \|\boldsymbol{c}_i\|_1 \quad \text{s.t.} \quad \boldsymbol{y}_i = \boldsymbol{Y}\boldsymbol{c}_i, \ c_{ii} = 0$$

solve the sparse optimization

$$\min \|\boldsymbol{c}_i\|_1 \quad \text{s.t.} \quad \boldsymbol{y}_i = \boldsymbol{Y}\boldsymbol{c}_i, \ c_{ii} = 0$$

$$\boldsymbol{c}_i = \begin{bmatrix} c_{i1} \\ c_{i2} \\ \vdots \\ c_{iN} \end{bmatrix}$$

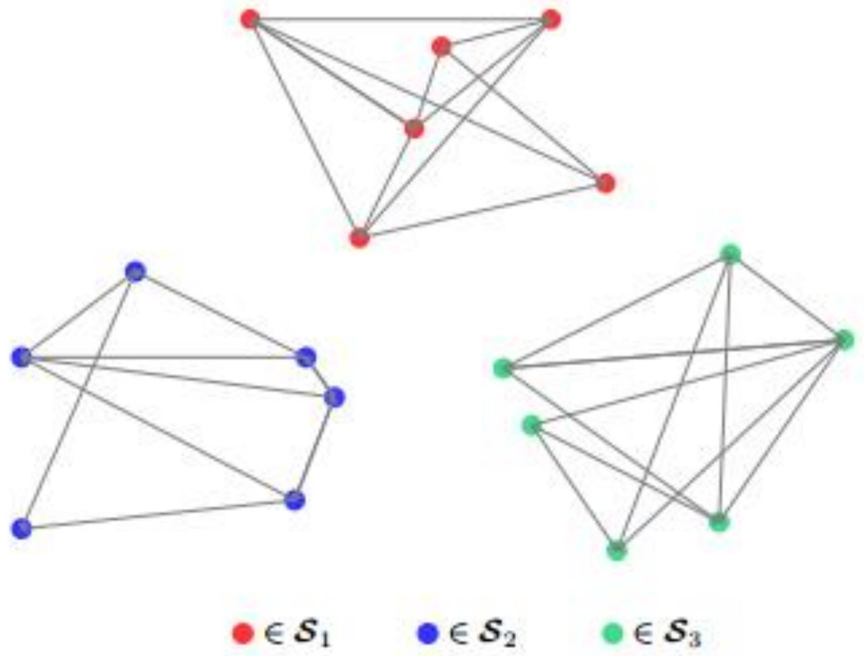
SSC algorithm

- 1: solve the sparse optimization

$$\min \|\boldsymbol{c}_i\|_1 \quad \text{s.t.} \quad \boldsymbol{y}_i = \boldsymbol{Y}\boldsymbol{c}_i,\; c_{ii} = 0$$

$$\boldsymbol{c}_i = \begin{bmatrix} c_{i1} \\ c_{i2} \\ \vdots \\ c_{iN} \end{bmatrix}$$

- 2: infer clustering from similarity graph
  - connect points using sparse weights
  - symmetrize the weights $w_{ij} = c_{ij} + c_{ji}$
  - apply spectral clustering



Subspace clustering

- All vertices representing data points in the same subspace form a **connected component** in the graph G = (V,E) where vertices V are the N data points and there is an edge $(v_i, v_j) \in \boldsymbol{E}$ when $C_{ji} \neq 0$
- In case of n-subspaces

$$\tilde{C} = \begin{bmatrix} \tilde{C}_1 & 0 & \cdots & 0 \\ 0 & \tilde{C}_2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \tilde{C}_n \end{bmatrix} \Gamma$$

- a new graph $\tilde{\boldsymbol{G}}$ with the adjacency matrix $\tilde{C}$ where $\tilde{C}_{ij} = |C_{ij}| + |C_{ji}|$

Subspace clustering

- Laplacian matrix of

  $\tilde{\mathbf{G}}$ is then formed by $L = D - \tilde{C}$ where $D \in \mathbb{R}^{N \times N}$ is a diagonal matrix with $D_{ii} = \sum_j \tilde{C}_{ij}$.

- Result from graph theory:-

  The multiplicity of the zero eigenvalue of the Laplacian matrix L corresponding to the graph $\tilde{\mathbf{G}}$ is equal to the number of connected components of the graph. Also, the components of the graph can be determined from the eigenspace of the zero eigenvalue. More

- Segmentation of data by applying k-means to a subset of eigenvectors of the Laplacian

Subspace clustering

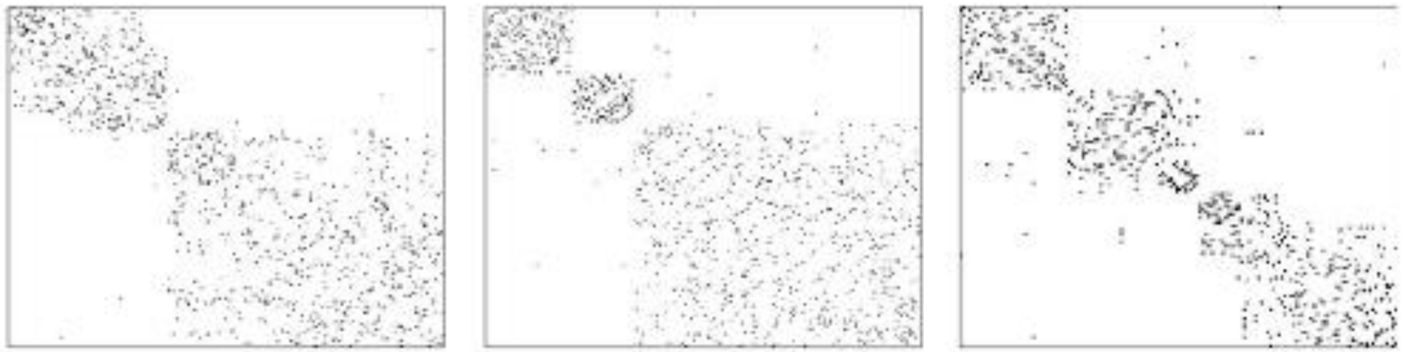


Figure 1. Sparse coefficients used to define the graph similarity matrix for three sequences: 1R2TRCT-g12, cars9, and articulated.

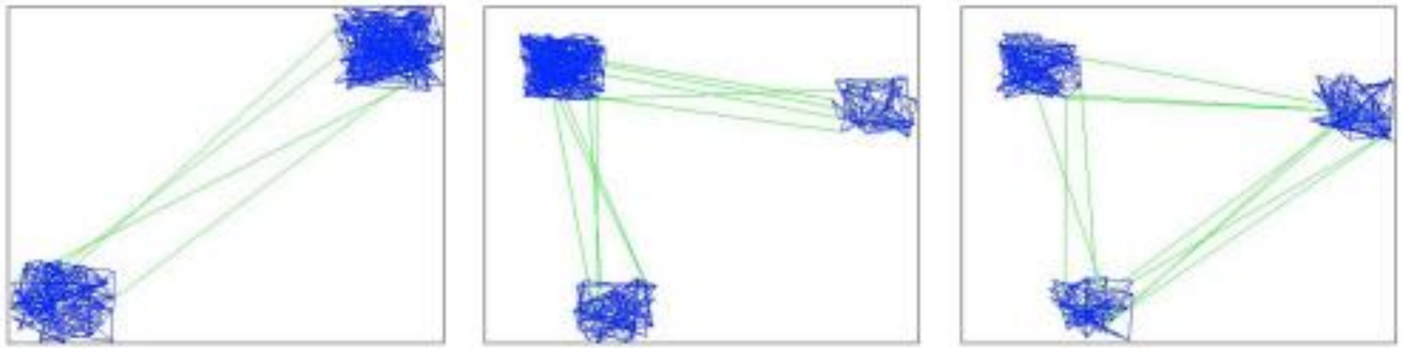


Figure 2. Similarity graphs for three sequences: 1R2TRCT-g12, cars9, and articulated.

Subspace clustering

- Similar extension for affine subspaces
- For noisy data (noise level bounded by ϵ) :-

$$\min \|\boldsymbol{c}_i\|_1 \quad \text{subject to} \quad \|Y_{\hat{i}}\boldsymbol{c}_i - \bar{\boldsymbol{y}}_i\|_2 \leq \epsilon.$$

- For noisy data (noise level unknown)

$$\min \|\boldsymbol{c}_i\|_1 + \gamma \|Y_{\hat{i}}\boldsymbol{c}_i - \bar{\boldsymbol{y}}_i\|_2$$

- For missing or corrupted data
  - Very similar approach as “Inpainting”

Results: motion segmentation

- motion segmentation problem, we consider the Hopkins 155 dataset, which consists of 155 video sequences of 2 or 3 motions corresponding to 2 or 3 low-dimensional subspaces in each video

Table 1. Classification errors (%) for sequences with 2 motions

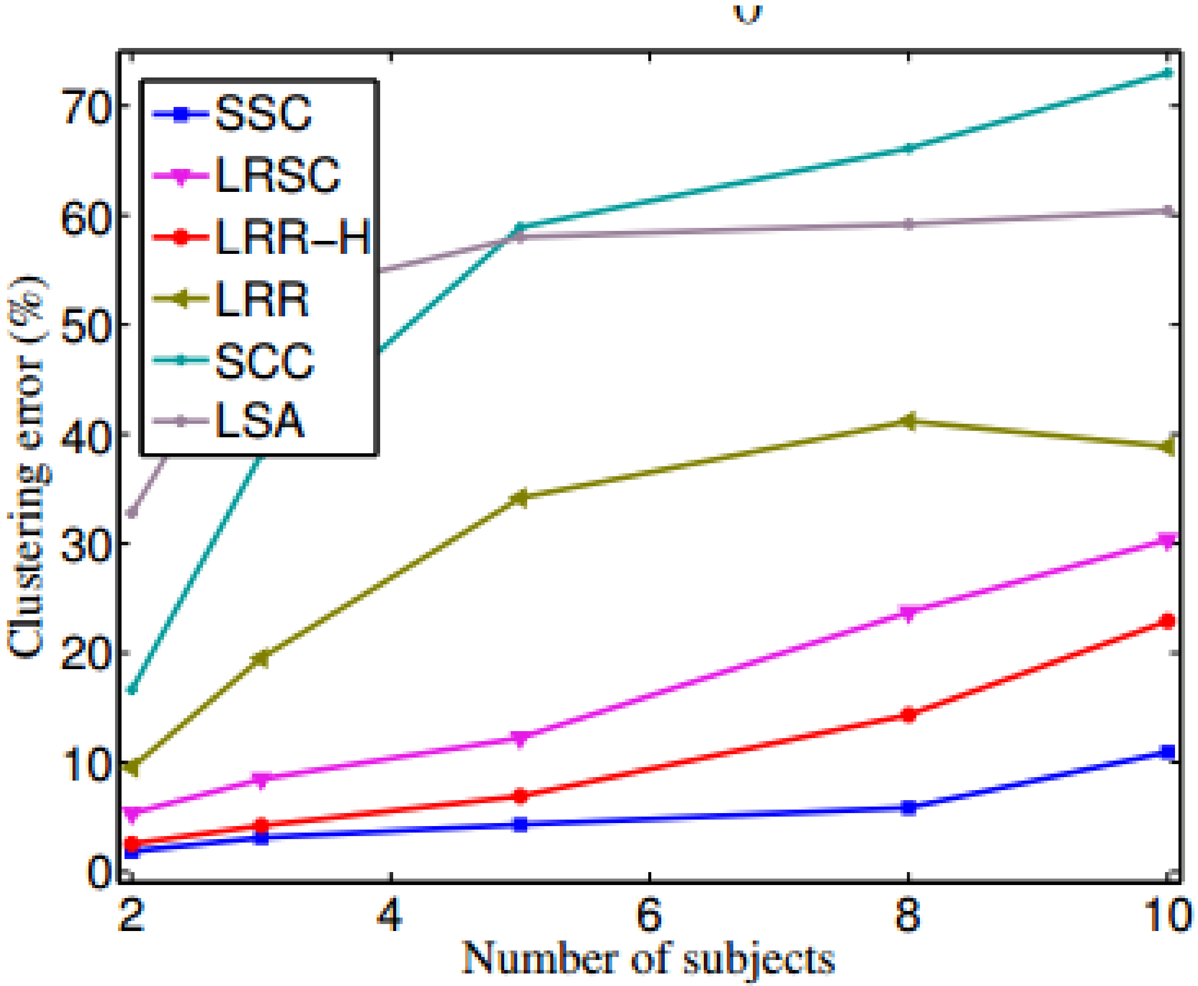
Method	GPCA	LSA	RANSAC	MSL	ALC	SSC-B	SSC-N
Checkerboard: 78 sequences							
Mean	6.09	2.57	6.52	4.46	1.55	**0.83**	1.12
Median	1.03	0.27	1.75	0.00	0.29	0.00	0.00
Traffic: 31 sequences							
Mean	1.41	5.43	2.55	2.23	1.59	0.23	**0.02**
Median	0.00	1.48	0.21	0.00	1.17	0.00	0.00
Articulated: 11 sequences							
Mean	2.88	4.10	7.25	7.23	10.70	1.63	**0.62**
Median	0.00	1.22	2.64	0.00	0.95	0.00	0.00
All: 120 sequences							
Mean	4.59	3.45	5.56	4.14	2.40	**0.75**	0.82
Median	0.38	0.59	1.18	0.00	0.43	0.00	0.00

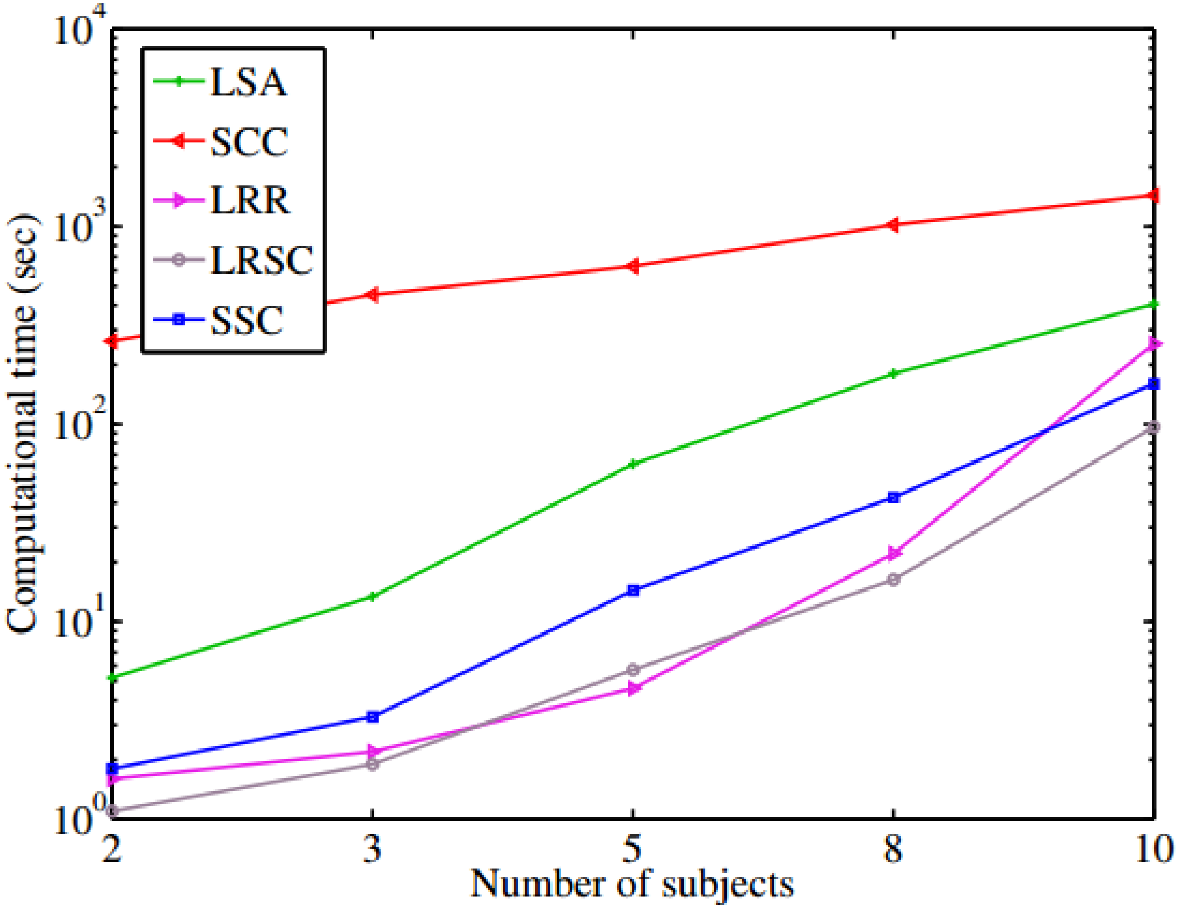
Table 2. Classification errors (%) for sequences with 3 motions

Method	GPCA	LSA	RANSAC	MSL	ALC	SSC-B	SSC-N
Checkerboard: 26 sequences							
Mean	31.95	5.80	25.78	10.38	5.20	4.49	**2.97**
Median	32.93	1.77	26.00	4.61	0.67	0.54	0.27
Traffic: 7 sequences							
Mean	19.83	25.07	12.83	1.80	7.75	0.61	**0.58**
Median	19.55	23.79	11.45	0.00	0.49	0.00	0.00
Articulated: 2 sequences							
Mean	16.85	7.25	21.38	2.71	21.08	1.60	**1.42**
Median	16.85	7.25	21.38	2.71	21.08	1.60	0.00
All: 35 sequences							
Mean	28.66	9.73	22.94	8.23	6.69	3.55	**2.45**
Median	28.26	2.33	22.03	1.76	0.67	0.25	0.20

Results: face clustering

- Ext YaleB faces





Sparse Subspace clustering: Claims

- Global sparse optimization
- Can deal with data points near the intersections
- Can deal with noise, outlying / missing entries
- Don’t require dimension / number of subspaces

Achieves/outperforms state-of-the-art results in

- segmentation of rigid-body motions
- clustering of face images
- temporal segmentation of videos

References

1. E. Elhamifar and R. Vidal, "Sparse subspace clustering," *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, Miami, FL, 2009, pp. 2790-2797. doi: 10.1109/CVPR.2009.5206547
2. E. Elhamifar and R. Vidal, "Sparse Subspace Clustering: Algorithm, Theory, and Applications," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765-2781, Nov. 2013. doi: 10.1109/TPAMI.2013.57
3. http://www.ccs.neu.edu/home/eelhami/cvpr15tutorial_files/Elhamifar_presentation_cvpr15.pdf
4. http://cis.jhu.edu/~rvidal/publications/SPM-Tutorial-Final.pdf
5. http://www.math.umn.edu/~lerman/Meetings/SIAM12_Ehsan.pdf
6. http://arxiv.org/pdf/1203.1005.pdf


Thanks!