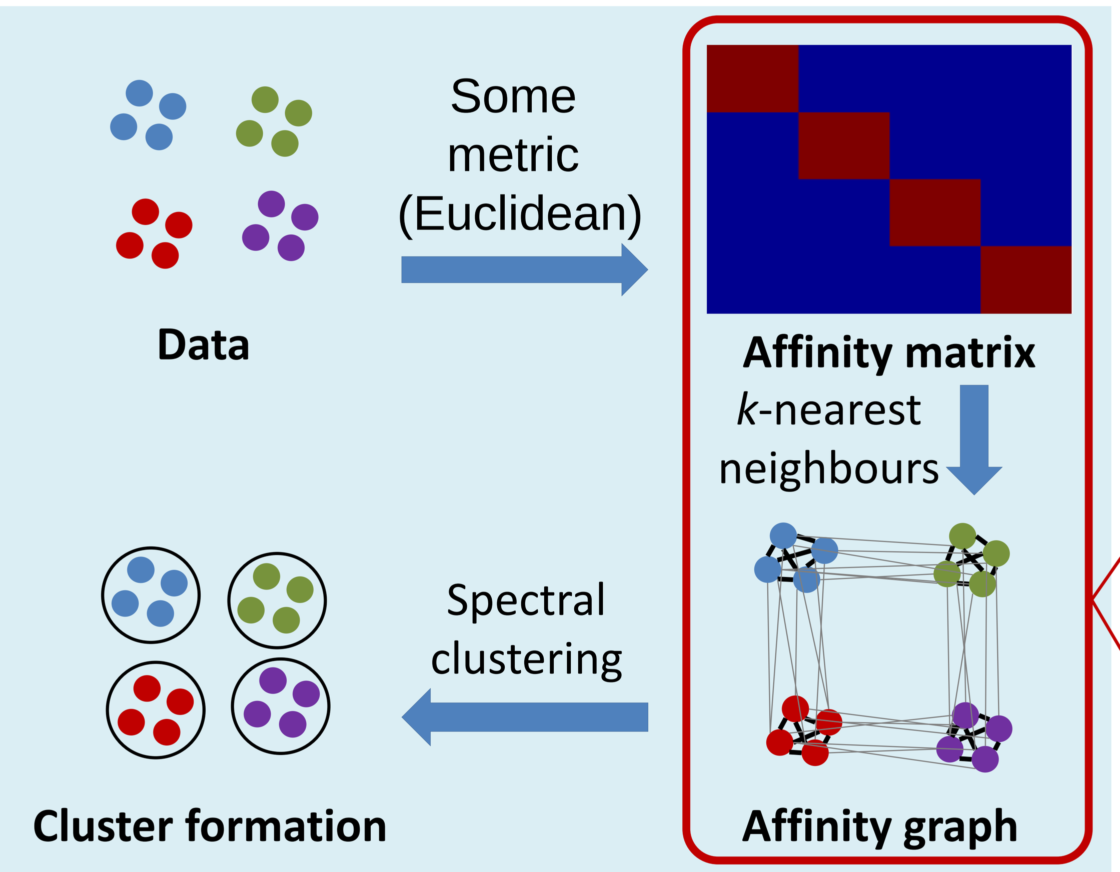
Constructing Robust Affinity Graphs for Spectral Clustering

Xiatian Zhu
Queen Mary University of London
xiatian.zhu@qmul.ac.uk

Chen Change Loy
The Chinese University of Hong Kong
ccloy@ie.cuhk.edu.hk

Shaogang Gong
Queen Mary University of London
s.gong@qmul.ac.uk

1 Problem Definition



Typical pipeline of spectral clustering

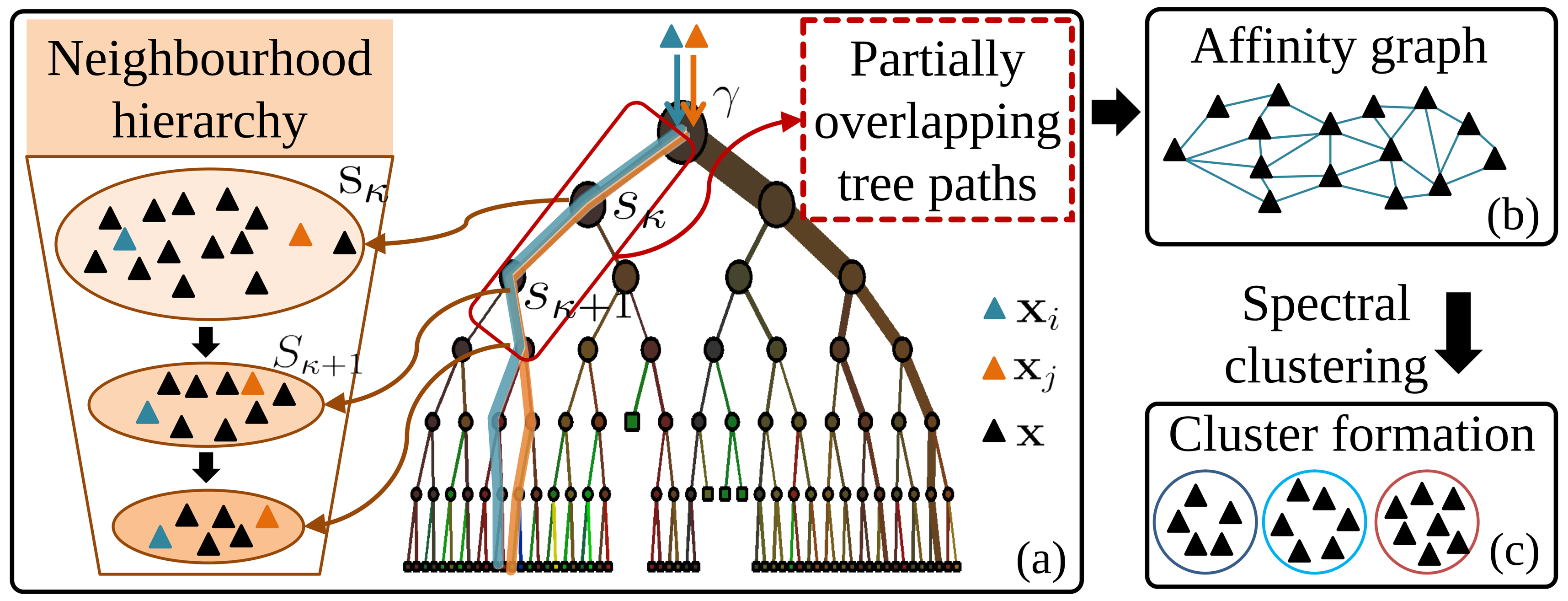
How to construct **robust** & **meaningful** affinity matrices/graphs?

Challenging data:
- High-dimensional
- Heterogeneous
- Noisy

Existing methods:
- the Euclidean metric + a Gaussian kernel to enforce locality [1]
- an adaptive scaling factor for the Gaussian kernel [2]
- random forest-based affinity graph construction [3,4,5]

Contributions: a generalised data similarity inference framework
- ✓ measure similarity via discriminative feature subspaces
- ✓ well motivated by information-theoretic definition of data similarity
- ✓ affinity matrix automatically possesses local neighbourhood

2 Structure-Aware Affinity Graphs



Merits of our model:

1. **Unsupervised**: based on clustering random forests
2. **Robust to noisy/irrelevant features**: define data pairwise similarity in discriminative feature subspaces
3. **Sense subtle similarities**: cumulate the weak pairwise proximities distributed over the entire tree hierarchies

Notations:

Tree paths of $\mathbf{x}_i$ & $\mathbf{x}_j$:
$$\mathcal{P}^i = \{\gamma, s_1^i, \ldots, s_\kappa^i, \ldots, \ell^i\}$$
$$\mathcal{P}^j = \{\gamma, s_1^j, \ldots, s_\kappa^j, \ldots, \ell^j\}$$

Path overlap of length λ:
$$\begin{cases} s_\kappa^i = s_\kappa^j, & \text{if } \kappa = \{1, \ldots, \lambda\}, \\ s_\kappa^i \neq s_\kappa^j, & \text{if } \kappa = \{\lambda+1, \ldots\}, \\ \ell^i \neq \ell^j. \end{cases}$$

- T_{clust}: tree number
- γ : root node
- s : split node
- $\mathcal{S}$: samples in s
- ℓ : leaf node
- Λ : samples in ℓ

The generalised ClustRF-Strct model

(1) Define the tree-level data pairwise similarity as:
$$a_{i,j}^t = \frac{\sum_{\kappa=1}^{\lambda} w_\kappa}{\sum_{\kappa=1}^{M} w_\kappa} \quad \text{where} \quad M = \max(|\mathcal{P}^i|, |\mathcal{P}^j|) - 1$$
w_κ: the tree node weight

(2) Construct tree-level affinity matrix A^t with elements as $a_{i,j}^t$

(3) Obtain the final affinity matrix: $A = \frac{1}{T_{\text{clust}}} \sum_{t=1}^{T_{\text{clust}}} A^t$

With different $a_{i,j}^t$, one can define a distinct affinity model

Variant I – the binary affinity model [3,4,5]

Idea: completely overlapped tree paths suggest strong data similarity:
$$w_\kappa = \begin{cases} 0, & \text{split node} \\ 1, & \text{leaf node.} \end{cases} \Rightarrow a_{i,j}^t = \begin{cases} 0, & \ell^i \neq \ell^j \\ 1, & \text{otherwise.} \end{cases}$$
Limitation: the weak knowledge in partially overlapped paths is ignored

Variant II – the uniform structure model

Idea: uniformly weight tree nodes to cumulate subtle similarities in partially overlapped paths, i.e.
$$w_\kappa = 1 \Rightarrow a_{i,j}^t = \frac{\lambda}{\max(|\mathcal{P}^i|, |\mathcal{P}^j|) - 1}$$
Limitation: distinct nodes may be not equally important as they reside at different tree layers and have data of dissimilarly complex distributions

Variant III – the adaptive structure model

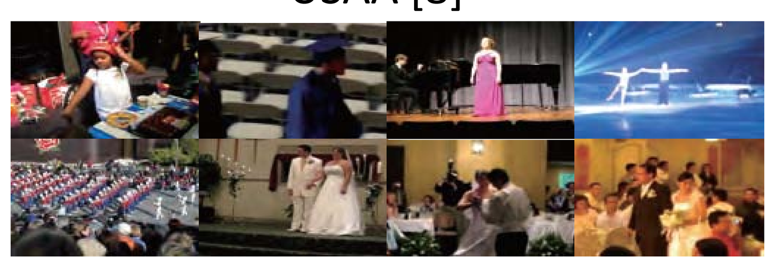
Idea: propose the hierarchical neighbourhood formed in clustering trees, which generalises the notion of adaptive nearest neighbours [6], formally,
$$w_\kappa = \frac{1}{|S_\kappa|} \Rightarrow a_{i,j}^t = \frac{\sum_{\kappa=1}^{\lambda}\left(\frac{1}{|S_\kappa|}\right)}{\sum_m \left(\frac{1}{|S_m^{\hat{b}}|}\right) + \frac{1}{|\Lambda^{\hat{b}}|}} \quad \text{where}$$
$$\hat{b} = \text{argmax}_{b \in \{i,j\}} |\mathcal{P}^b|$$

3 Evaluations

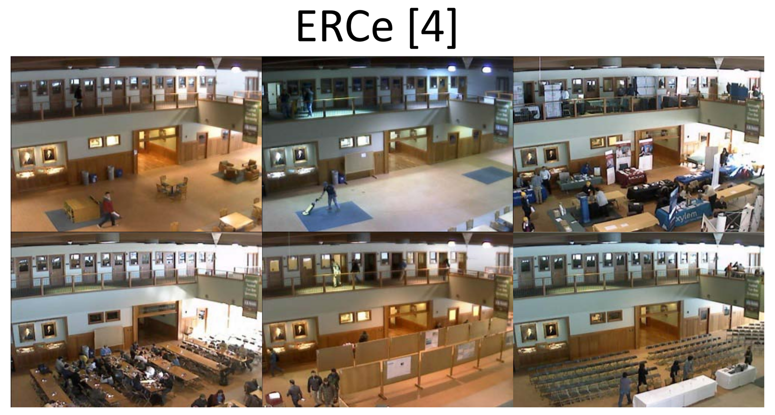
Datasets

CMU-PIE [7]

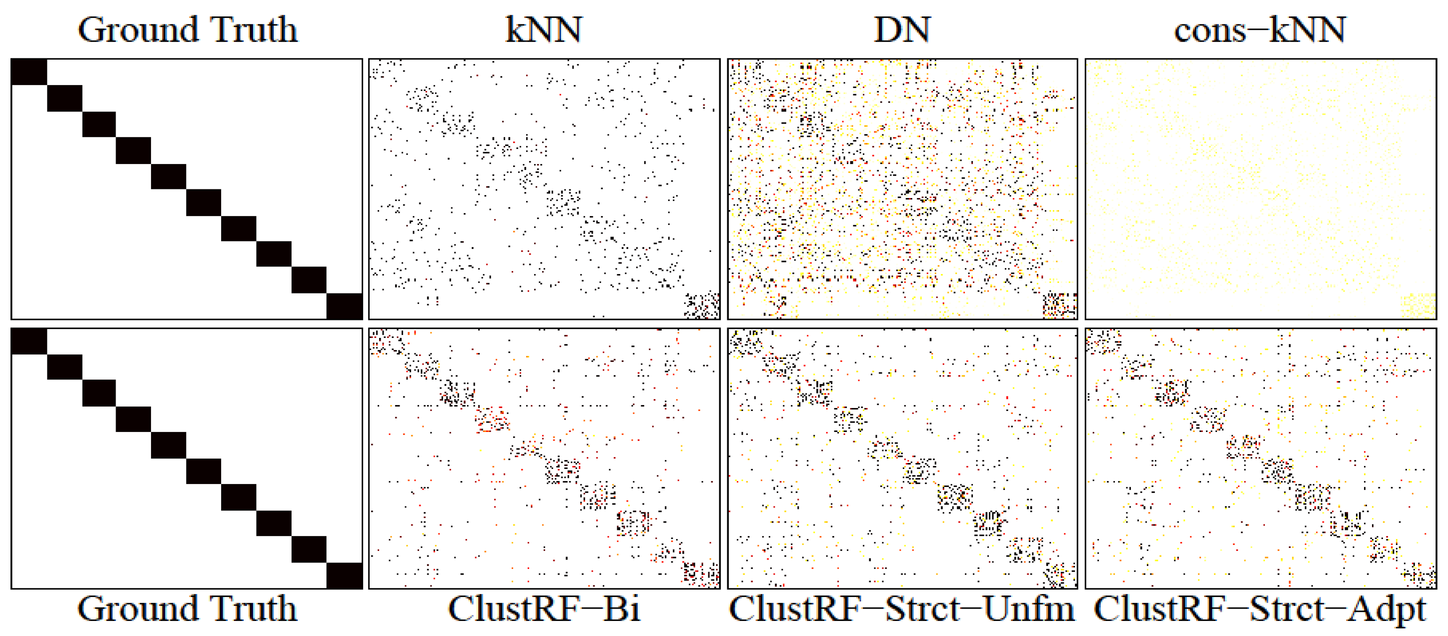
USAA [8]



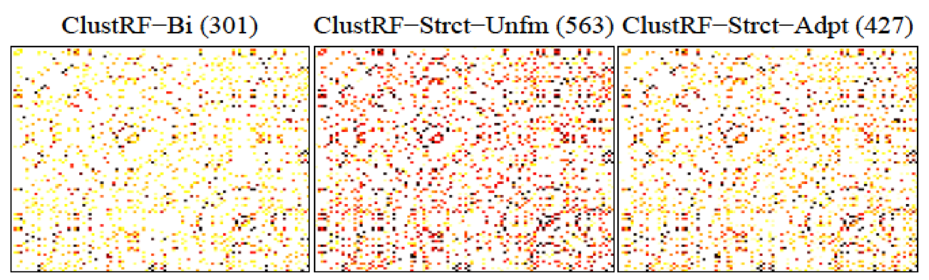
ERCe [4]



Evaluation of affinity matrices



The affinity matrices induced by different approaches on CMU-PIE



Compare forest based models: the affinity between face images from the same person in CMU-PIE. (X): higher X is better.

Evaluation of clustering performance

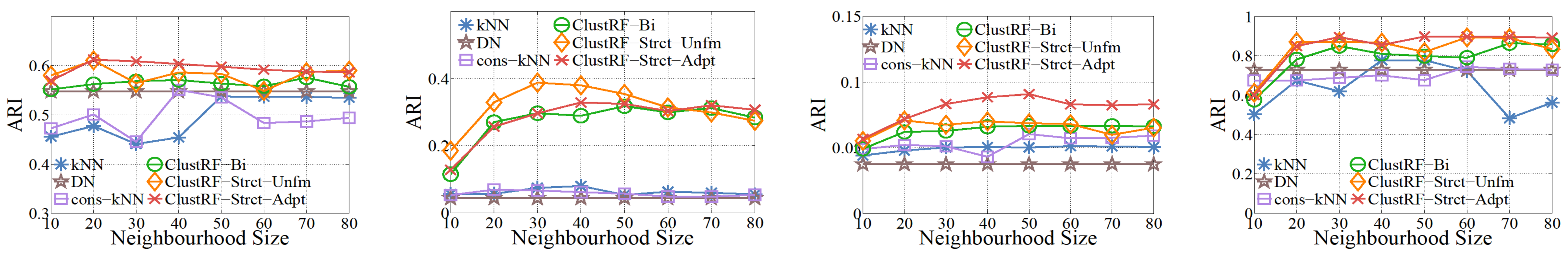


Image Segmentation [9] | CMU-PIE [7] | USAA [8] | ERCe [4]

Compare adjusted Rand Index (ARI) curves over different neighbourhood sizes. Higher is better.

Dataset	Image Segmentation [9]					CMU-PIE [7]					USAA [8]					ERCe [4]				
M	20	40	60	80	100	20	40	60	80	100	20	40	60	80	100	20	40	60	80	100
*k*NN[1]	34.8	36.2	37.6	37.8	37.9	4.4	4.4	4.9	4.8	4.7	3.5	3.1	3.3	3.6	3.6	45.9	48.1	52.1	52.7	51.8
DN[10]	38.3	29.1	34.7	37.2	37.2	3.0	2.3	2.4	3.0	3.5	2.6	2.3	2.5	2.0	1.7	51.0	52.1	49.9	18.3	25.6
Cons-*k*NN[11]	34.9	36.8	35.8	36.8	35.9	4.0	4.4	4.3	4.3	4.2	3.8	3.8	3.8	3.8	3.9	49.2	52.1	52.0	52.0	55.7
ClusterRF-Bi[3,4,5]	39.5					19.8					4.5					56.1				
ClustRF-Strct-Unfm	40.7					**22.9**					4.7					59.3				
ClustRF-Strct-Adpt	**41.8**					20.5					**5.7**					**60.4**				

Compare area under ARI curves. Higher is better. For all Euclidean distance based models, we vary the Gaussian kernel scale (i.e. varying M) used for converting the distance matrix into the affinity matrix

[1] J. Wang, et al. CVPR'08
[2] L. Zelnik-manor and P. Perona. NIPS'04
[3] A. Criminisi et al. FTCGV'12
[4] X. Zhu, et al. ICCV' 13
[5] Y. Pei et al. ICCV' 13
[6] Y. Lin and Y. Jeon. J AM STAT ASSOC'06
[7]T. Sim et al. TPAMI' 03
[8] Y. Fu et al. TPAMI' 13
[9] UCI repository. 2007
[10] M. Pavan and M. Pelillo. TPAMI' 07
[11] V. Premachandran and R. Kakarala. CVPR' 13

Code available: http://www.eecs.qmul.ac.uk/~xz303/