Agglomerative Hierarchical Kernel Spectral Data Clustering (AH-KSC)

Raghvendra Mall[1], Rocco Langone[1],
Johan A.K. Suykens[1]

[1]KU Leuven

IEEE SSCI CIDM 2014

Introduction

- Clustering is widely used technique in data mining and machine learning [1].
- Core concept is to identify inherent groups present within the data.
- Real-life data exhibits hierarchical clustering organization [2].
- Spectral clustering methods [3, 4, 5] are state-of-the-art.
- Perform eigen-decomposition of a pairwise similarity matrix resulting in localized structures for the k clusters in the eigenspace.
- Computationally infeasible $O(N^3)$ and memory extensive $O(N^2)$ for large data.
- Provide flat clustering and not easy to extend it to obtain hierarchical clusters.

Background

- A kernel spectral clustering (KSC) [6] method based on weighted kernel PCA (wk-PCA) was formulated in a primal-dual setting.
- The KSC learning framework has a training, validation and test phase.
- The model parameters like k (number of clusters) and kernel parameter σ/σ_{χ^2} are obtained during validation stage.
- KSC model has a powerful out-of-sample extensions property which allows cluster affiliation for previously unseen data.

Related Work

- A hierarchical kernel spectral clustering (HKSC) was proposed in [7].
- Multiple scales of σ were used to generate KSC model for different k resulting in independent cluster memberships at various levels of hierarchy.
- Special linkage criterion [7] was designed to merge these clusters.
- One drawback is that as cluster memberships are obtained from different KSC models for different values of k, σ, the merge is usually not perfect.
- Several points are forced to merge using the majority rule [7].
- An Agglomerative Hierarchical KSC (AH-KSC) was proposed in [8, 9] for networks which overcomes the shortcomings of HKSC.

Goal

Given a dataset detect multiples levels of hierarchy with meaningful clusters at each level of hierarchy using a modified version of Agglomerative Hierarchical Kernel Spectral Clustering (AH-KSC) technique [8, 9] oriented for data and images.

Real-Image Illustration

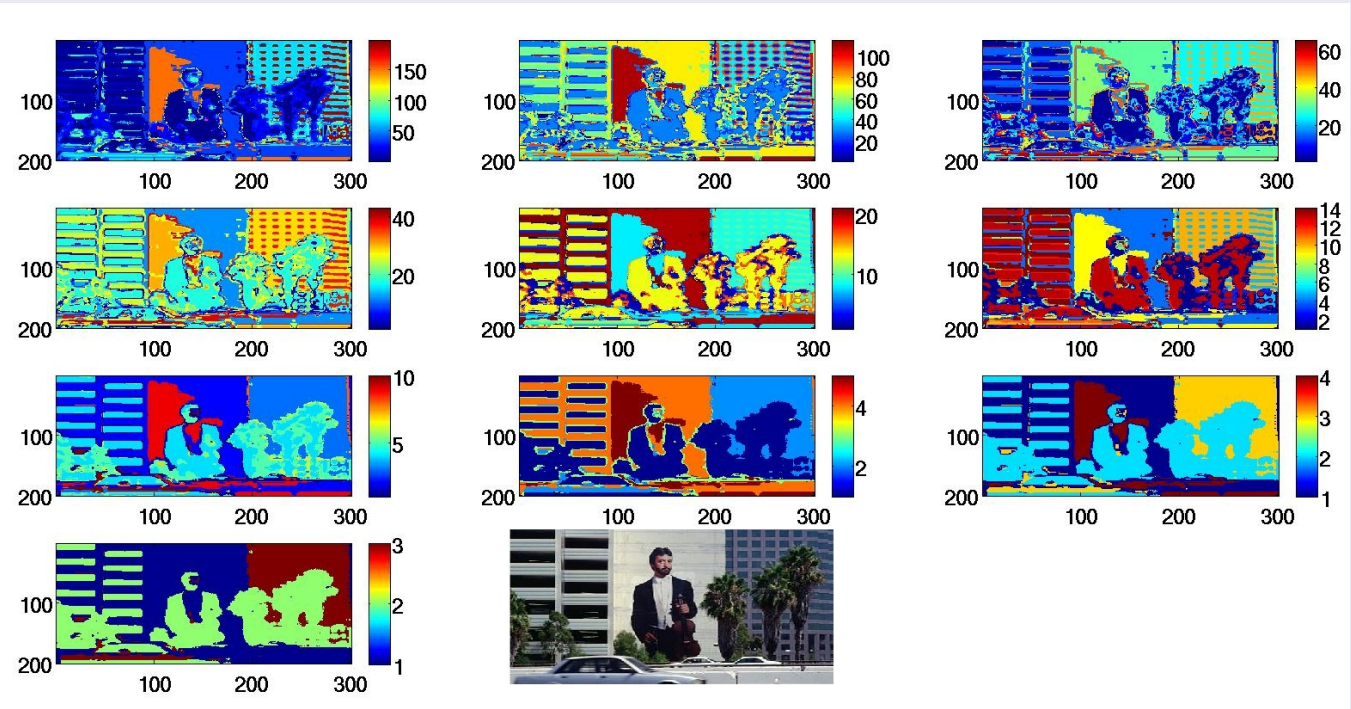


Figure: Result of AH-KSC method on a real world image dataset.

Contributions

- Extend AH-KSC technique from complex networks to datasets and images.
- Obtain optimal model parameters (σ, k) using balanced angular fitting (BAF) criterion [10] which generally corresponds to one level of hierarchy.
- BAF criterion can capture multiple peaks corresponding to different levels of hierarchy for an optimal σ.
- Exploit eigen-projections to obtain a set ($\mathcal{T}$) of distance thresholds.
- Utilize distance thresholds ($t_i \in \mathcal{T}$) to obtain multiple levels of hierarchy for datasets and images.

Building Blocks

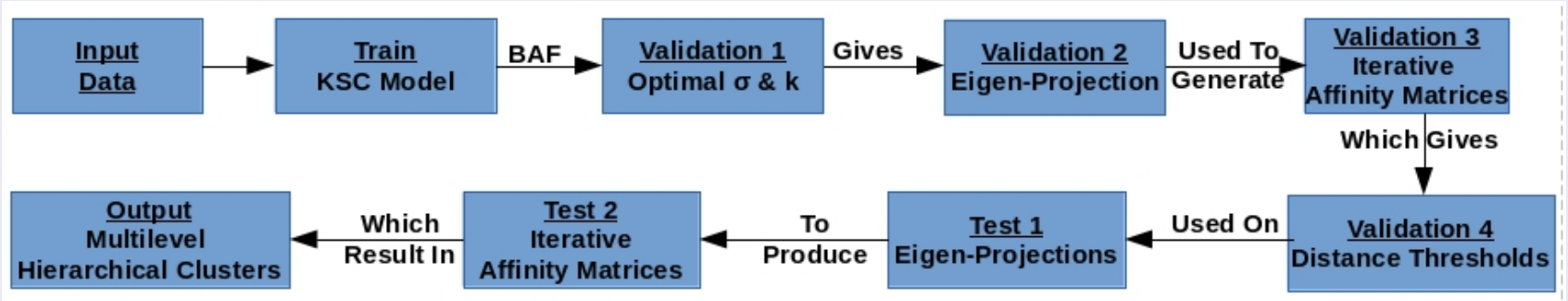


Figure: Steps of AH-KSC method [8, 9] with additional step where the optimal σ and k are estimated.

Kernel Spectral Clustering

KSC Formulation

Given $\mathcal{D} = \{x_i\}_{i=1}^{N_{tr}}$ and *maxk*, the primal formulation of the weighted kPCA is given by:

$$\min_{w^{(l)}, e^{(l)}, b_l} \quad \frac{1}{2} \sum_{l=1}^{maxk-1} w^{(l)\intercal} w^{(l)} - \frac{1}{2N_{tr}} \sum_{l=1}^{maxk-1} \gamma_l e^{(l)\intercal} D_\Omega^{-1} e^{(l)} \tag{1}$$

$$\text{such that} \quad e^{(l)} = \Phi w^{(l)} + b_l 1_{N_{tr}}, l = 1, \ldots, maxk - 1$$

KSC Primal-Dual Model

- Primal clustering model is given as: $e_i^{(l)} = w^{(l)\intercal} \phi(x_i) + b_l, i = 1, \ldots, N_{tr}$.
- Corresponding dual problem becomes: $D_\Omega^{-1} M_D \Omega \alpha^{(l)} = \lambda_l \alpha^{(l)}$.
- Dual predictive model is given as: $\hat{e}^{(l)}(x) = \sum_{i=1}^{N_{tr}} \alpha_i^{(l)} K(x, x_i) + b_l$.
- Each element of Ω, denoted as $\Omega_{ij} = K(x_i, x_j) = \phi(x_i)^\intercal \phi(x_j)$ is obtained for example using radial-basis function (RBF) kernel.

Model Selection

Balanced Angular Fitting

- The Balanced Angular Fitting (BAF) criterion [10] is defined as:

$$\begin{aligned} &\text{BAF}(k,\sigma) = \eta \sum_{p=1}^{k} \sum_{valid_{(i,\sigma)} \in Q_p} \frac{1}{k} \cdot \frac{MS(valid_{(i,\sigma)})}{|Q_p|} + (1-\eta)\frac{\min_l |Q_l|}{\max_m |Q_m|}, \\ &MS(valid_{(i,\sigma)}) = \max_j \cos(\theta_{j,valid_{(i,\sigma)}}), j = 1,\ldots,k \\ &cos(\theta_{j,valid_{(i,\sigma)}}) = \frac{\mu_j^{\mathsf{T}} e_{valid_{(i,\sigma)}}}{\|\mu_j\| \|e_{valid_{(i,\sigma)}}\|}, j = 1,\ldots,k. \end{aligned} \tag{2}$$

- Q_p represents the set of validation points belonging to cluster p.
- BAF criterion is based on angular similarity of validation points to cluster means.
- Validation points are allocated to clusters (μ_j) based on least angular distance.
- BAF criterion varies from $[-1, 1]$ and higher values are better for a given k.

Model Selection

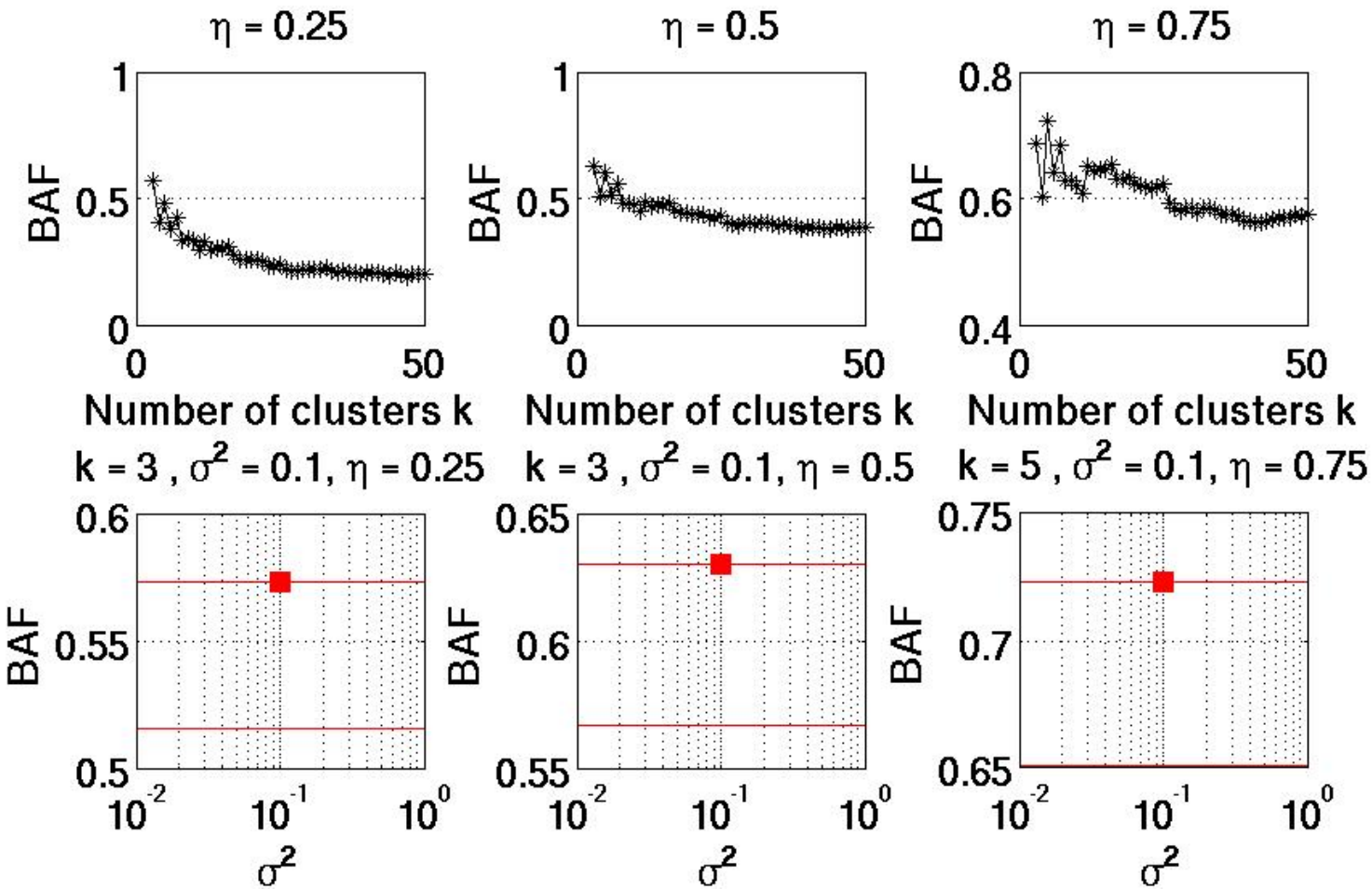


Figure: Selection of optimal σ using BAF criterion and illustration of multiple peaks in the BAF curve for different values of k for image dataset.

AH-KSC method

Determining Distance Thresholds

1. Create $S^{(0)}_{valid}$ for the validation projections (P_{valid}) as:
$$S^{(0)}_{valid}(i,j) = CosDist(e_i, e_j) = 1 - \cos(e_i, e_j) = 1 - \frac{e_i^{\mathsf{T}} e_j}{\|e_i\|\|e_j\|}.$$
2. Set initial distance threshold $t^{(0)} \in 0.15$ & $h = 0$.
3. *GreedyMaxOrder* [8, 9] gives $C^{(h)}$ and k at level h.
4. Generate affinity matrix at level h of hierarchy as:
$$S^{(h)}_{valid}(i,j) = \frac{\sum_{m \in C_i^{(h-1)}} \sum_{l \in C_j^{(h-1)}} S^{(h-1)}_{valid}(m,l)}{|C_i^{(h-1)}| \times |C_j^{(h-1)}|} \tag{3}$$
5. Threshold $t^{(h)} = \text{mean}(\min_j(S^{(h)}_{valid}(i,j))), i \neq j$ at level h.
6. Repeat Steps 4,5,6 till $k = 1$.
7. Obtain $\mathcal{T} = \{t^{(0)}, \ldots, t^{(maxh)}\}$.

AH-KSC method

Agglomerative Hierarchical KSC for Test Data

- Retrieve eigen-projections (P_{test}) for test set.
- To self-tune the approach use $t^{(i)} > t^{(0)}$, $i > 0$.
- Use *GreedyFirstOrder* [8, 9] to obtain $S_{test}^{(2)}$, $C^{(1)}$ and k.
- *For each* $t^{(h)} \in \mathcal{T}$, $h > 1$ do
  1. $[C^{(h)}, k] = GreedyMaxOrder(S_{test}^{(h)}, t^{(h)})$.
  2. Add $C^{(h)}$ to the set $\mathcal{C}$.
  3. Create $S_{test}^{(h+1)}$ using $S_{test}^{(h)}$ and $C^{(h)}$ as shown in (3).
- Results in the set $\mathcal{C}$ with cluster memberships.
- Propagate cluster memberships from lowest to coarsest level of hierarchy.

AH-KSC Illustration

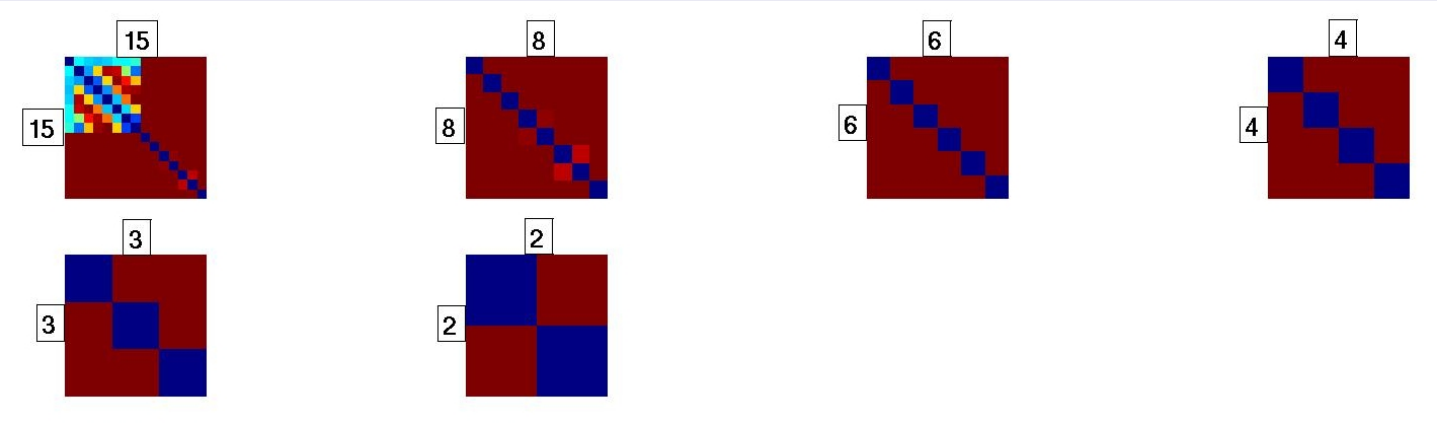


Figure: $S^{(h)}$ created at different levels of hierarchy in left to right order.

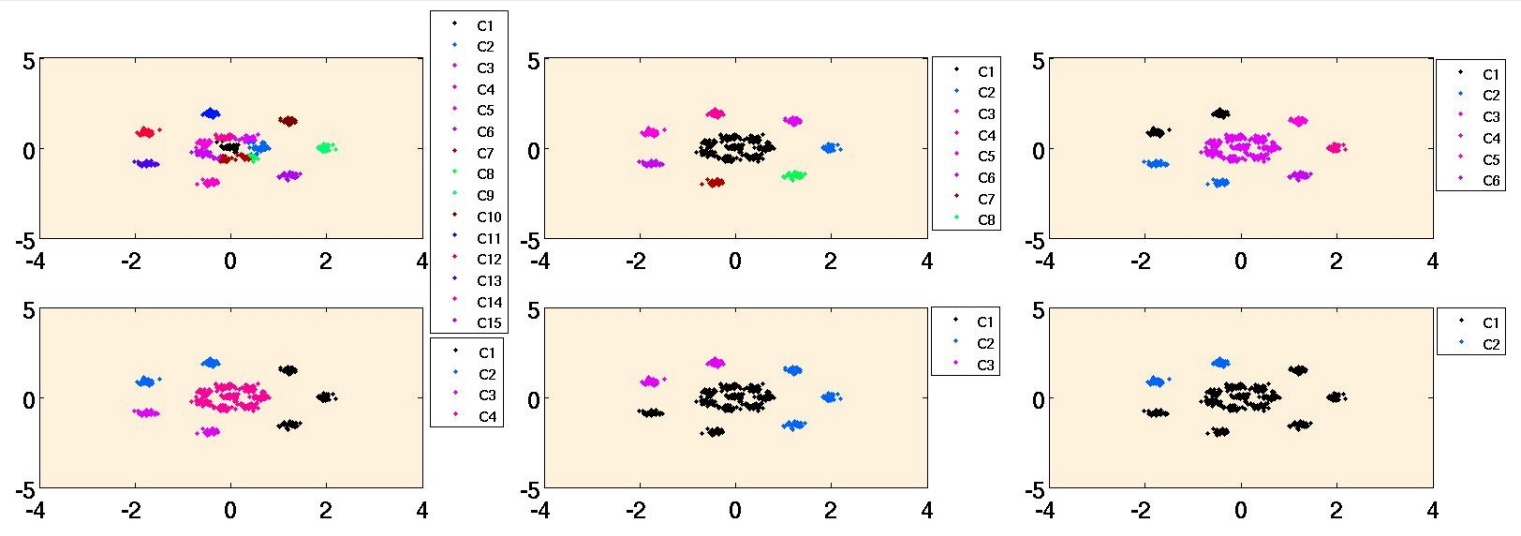


Figure: The cluster memberships at different levels of hierarchy.

Experiments

Experimental Setup

- Datasets obtained from `http://cs.joensuu.fi/sipu/datasets/` and images obtained from Berkeley image segmentation database.
- Use 30% and 40% of data as training and validation set.
- Compared on quality metrics like Silhouette (Sil), Davies-Bouldin (DB) [11] index and F-score defined as ratio between Sil and DB.
- Compared with single (SL), complete (CL), median (ML), weighted (WL) and average (AL) link hierarchical clustering techniques.

Datasets			
Name	Total Points (N)	Dimensions	Optimal k
Aggregation	788	2	7
Dim2	1351	2	9
R15	600	2	15
S1	5000	2	15
Images			
Number	Pixels		Optimal k
Id:10001	6,400		4
Id:10002	6,400		5
Id:100007	154,401		4
Id:119082	154,401		5

Table: For datasets optimal k is equal to ground truth number of clusters. For images optimal k is selected based on the maximum F-score value for SL, CL, ML, WL and AL agglomerative hierarchical clustering techniques.

Experiments

Dataset Results

Dataset	AH-KSC	SL	AL	CL	ML	WL
	F-score					
Aggregation	**1.34**	0.84	1.08	1.01	0.98	0.92
Dim2	2.27	**7.37**	5.30	4.54	5.40	5.40
R15	**4.75**	0.92	3.32	3.28	3.18	3.15
S1	**2.96**	0.0	1.81	1.55	1.58	1.76
	DB					
Aggregation	**0.466**	0.65	0.64	0.66	0.68	70
Dim2	0.37	**0.12**	0.17	0.20	0.17	0.17
R15	**0.17**	0.81	0.27	0.27	0.28	0.29
S1	**0.29**	1.7	0.45	0.51	0.50	0.46
	Sil					
Aggregation	**0.69**	0.55	**0.69**	0.67	0.66	0.65
Dim2	0.85	**0.90**	**0.90**	**0.90**	**0.90**	**0.90**
R15	0.78	0.74	**0.90**	0.90	0.90	0.90
S1	**0.88**	0.54	0.81	0.79	0.79	0.81
	ARI					
Aggregation	**0.94**	0.81	0.93	0.78	0.67	0.71
R15	0.87	0.54	**0.99**	0.98	0.98	0.99

Table: Comparison of AH-KSC method with single, average, complete, median and weight linkage techniques based on various quality measures. The highlighted result represent the best approach.

Aggregation Dataset Result

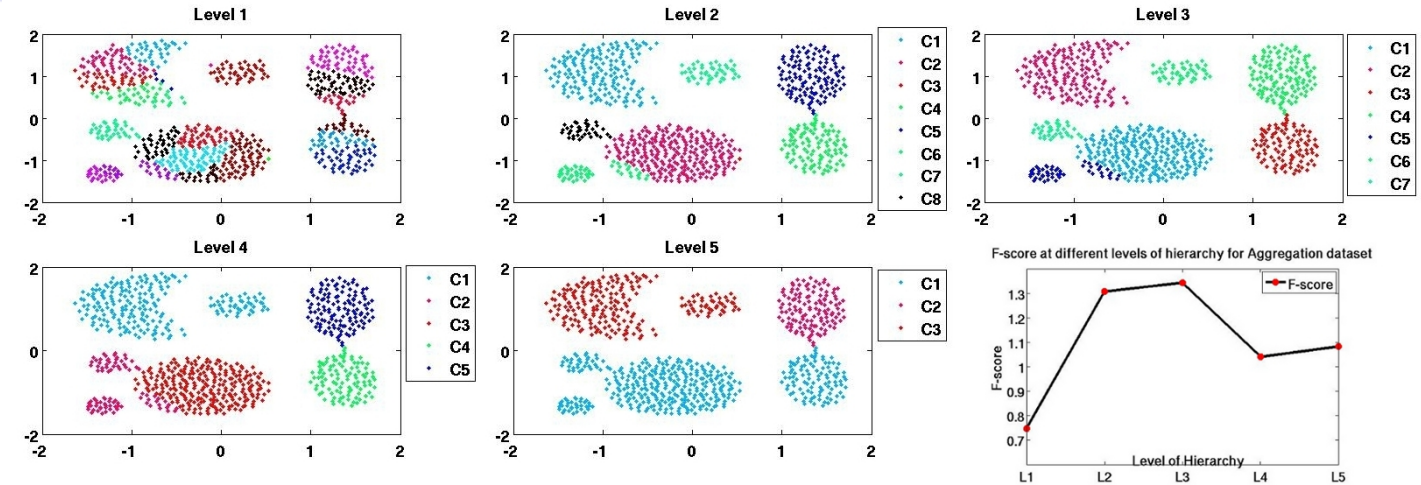


Figure: 5 levels of hierarchy produced by AH-KSC method and best F-score at $k = 7$ for Level 3.

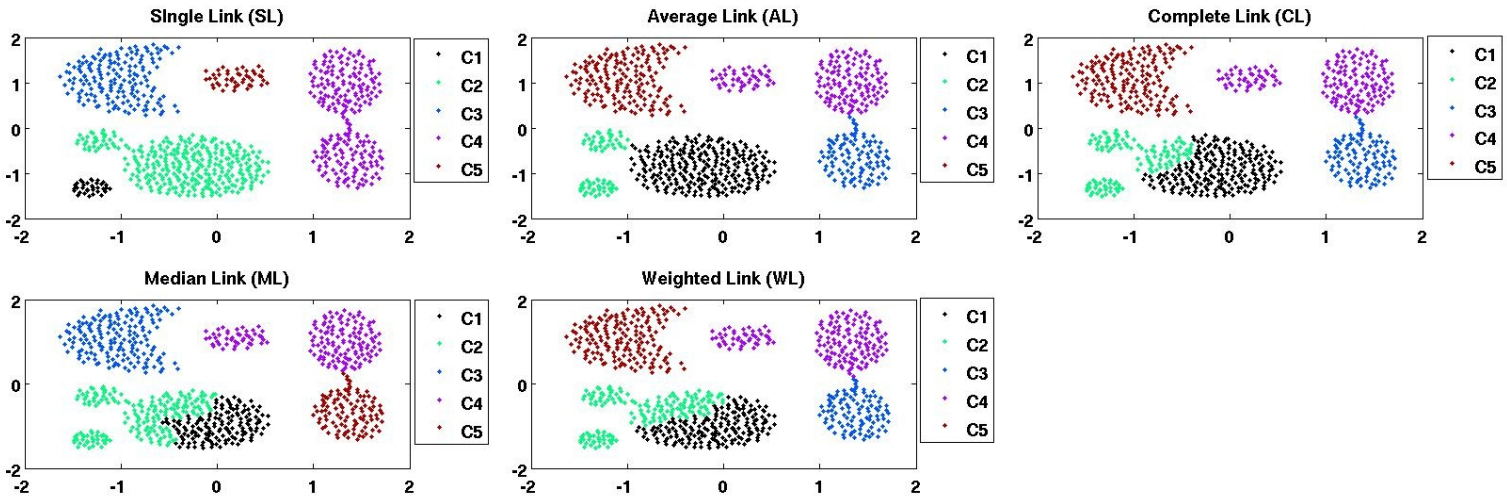


Figure: Best results for other hierarchical techniques where F-score is maximum for $k = 5$.

Image Results

Image	AH-KSC	SL	CL	ML	WL
	F-score				
Id:10001	**2.08**	0.394	1.57	1.35	1.35
Id:10002	1.02	0.70	1.37	1.2	**1.40**
Id:100007	0.73	0.38	0.41	0.68	**1.00**
Id:119082	**0.59**	0.57	0.44	0.56	0.59
	DB				
Id:10001	**0.41**	0.68	0.55	0.62	0.62
Id:10002	0.73	0.80	0.57	0.63	**0.55**
Id:100007	0.89	1.02	1.41	0.75	**0.48**
Id:119082	0.97	**0.80**	1.1	0.88	0.81
	Sil				
Id:10001	**0.85**	0.27	0.85	0.84	0.84
Id:10002	0.75	0.56	**0.77**	0.76	0.77
Id:100007	**0.65**	0.39	0.58	0.51	0.48
Id:119082	**0.57**	0.45	0.48	0.49	0.49

Table: Comparison of AH-KSC method with other agglomerative hierarchical techniques based on various quality measures for images. The highlighted result represent the best approach.

Image Segmentation Result

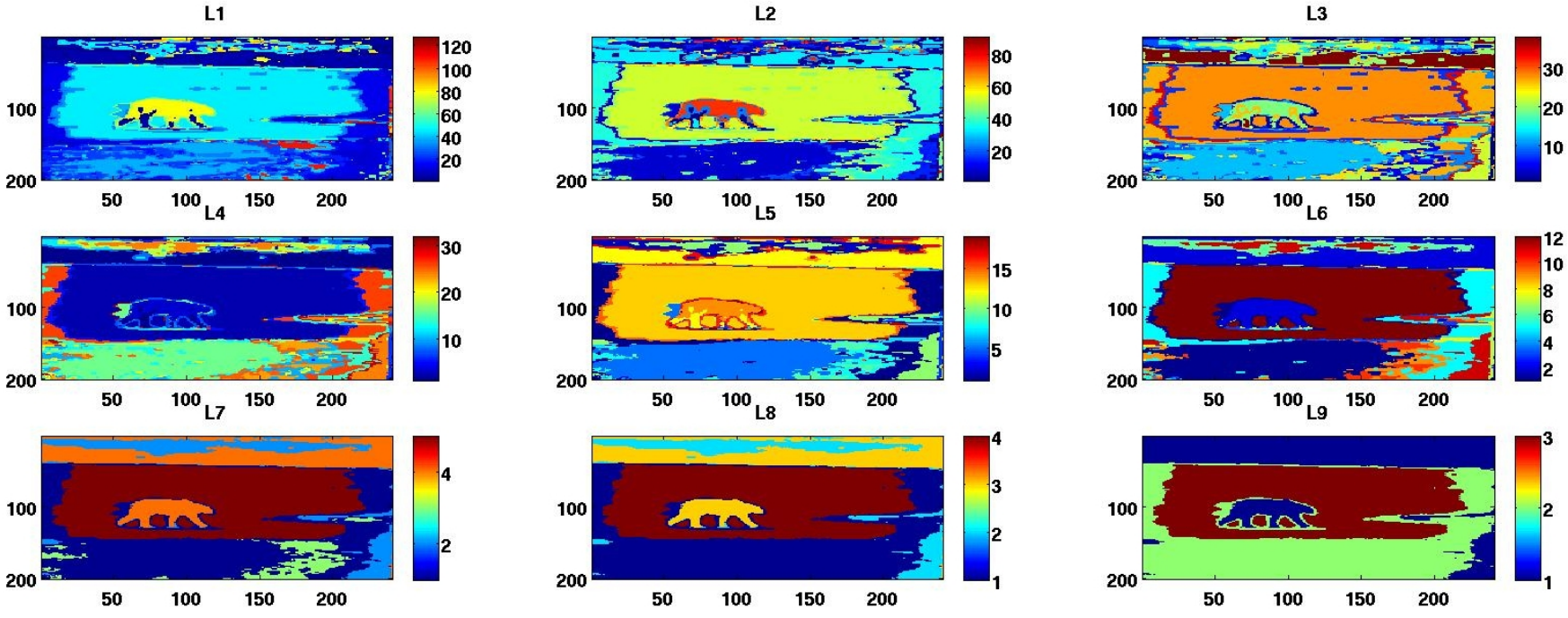
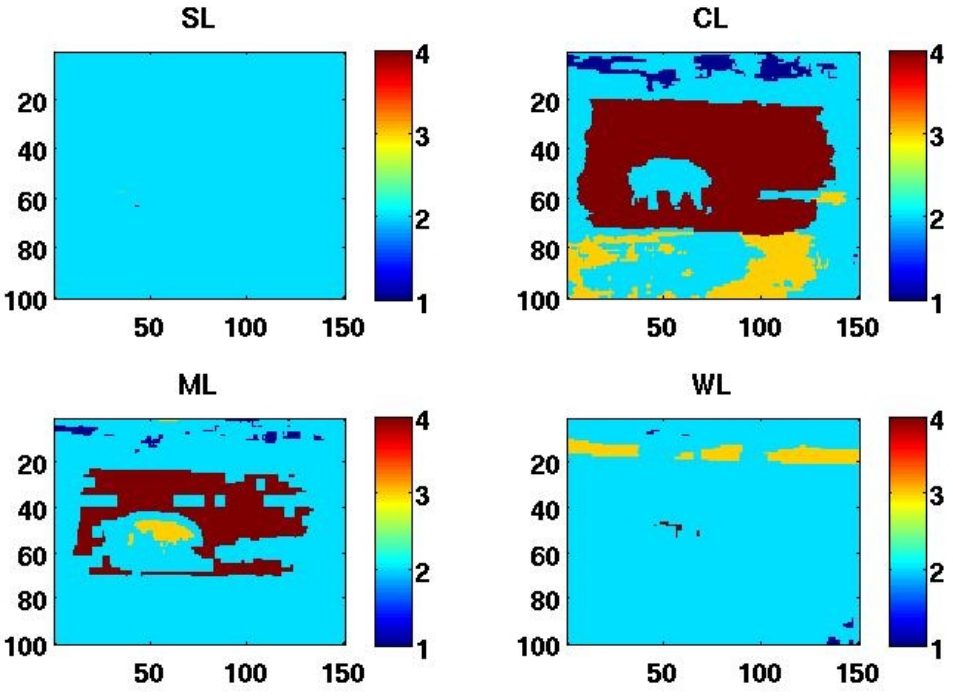
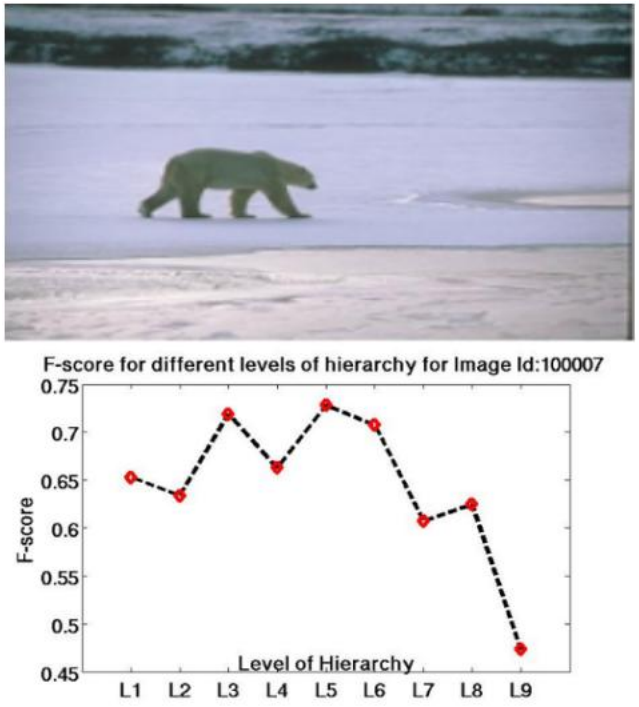


Figure: Hierarchical Segmentation by AH-KSC method from levels L1 to L9.

Image Segmentation Result



(a) Best results for other hierarchical techniques where F-score is maximum at $k = 4$.



(b) Original Image & Best F-score at L5 when $k = 19$ for AH-KSC method.

Figure: Hierarchical segmentation results by different methods on image Id:100007.

Conclusion

- Extended AH-KSC [8, 9] method from networks to datasets and images.
- Optimal kernel parameter σ selected using BAF criterion.
- Exploited the eigen-projections to obtain affinity matrices resulting in a set of increasing distance thresholds $\mathcal{T}$.
- Used these distance thresholds to perform agglomerative hierarchical clustering in bottom-up fashion.
- Compared AH-KSC method with several state-of-the-art agglomerative clustering techniques.



A.K. Jain and P. Flynn.

Image segmentation using clustering.
pages 65–83, 1996.



H. Yu, J. Yang, and J. Han.

Classifying large data sets using svms with hierarchical clusters.
In *Proc. of KDD*, pages 306–315, 2003.



A.Y. Ng, M.I. Jordan, and Y. Weiss.

On spectral clustering: analysis and an algorithm.
In *Proceedings of the Advances in Neural Information Processing Systems*, pages 849–856. MIT Press: Cambridge, MA, 2002.



J. Shi and J. Malik.

Normalized cuts and image segmentation.
IEEE Transactions on Pattern Analysis and Intelligence, 22(8):888–905, 2000.



U. von Luxburg.

A tutorial on spectral clustering.
Statistical Computing, 17:395–416, 2002.



C. Alzate and J.A.K. Suykens.

Multiway spectral clustering with out-of-sample extensions through weighted kernel pca.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(2):335–347, 2010.



C. Alzate and Johan A.K. Suykens.

Hierarchical kernel spectral clustering.
Neural Networks, 35:21–30, 2012.



R. Mall, R. Langone, and J.A.K. Suykens.

Multilevel hierarchical kernel spectral clustering for real-life large scale complex networks.
PLOS One, 9(6), 2014.


R. Mall, R. Langone, and J.A.K. Suykens.
Agglomerative hierarchical kernel spectral clustering for large scale networks.
In *Proc. of ESANN*, 2014.


R. Mall, R. Langone, and Johan A.K. Suykens.
Kernel spectral clustering for big data networks.
Entropy (Special Issue: Big Data), 15(5):1567–1586, 2013.


R.Rabbany, M.Takaffoli, J.Fagnan, O.R.Zaiane, and R.J.G.B.Campello.
Relative validity criteria for community mining algorithms.
In *Proc. of ASONAM*, pages 258–265, 2012.