

Segmentation of unsupervised multispectral satellite imaginary using mean shift, minimum spanning tree and watershed based clustering techniques

¹J.Senthil Murugan, ²A.Kavithamani, ³M.keerthana

¹Assistant Professor, ^{2,3}PG Scholar, Department of MCA

¹Senthilmuruganmtech@gmail.com, ²Kavimani923@gmial.com, ³keerthana.ranjani@gmail.com

Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College.

Avadi, Chennai-62.

Abstract- An unsubstantiated density centered subdivision; merging an altered mean-shift (MS) and an original tinniest traddling tree (TST) based gathering method of the slightest bit detected satellite metaphors has stayed projected in this communication. The image is first preprocessed by an adapted variety of the typical MS created dissection which conserves the necessary discontinuities contemporary in the twin and assurances completed separation in the productivity. An adjacent neighbor based technique for approximating the bandwidth of the grain mass estimator (GME) and a original closure condition have been unified into the customary MS. Since the segmented sections as bulges in a squat equal article interplanetary, a TST is assembled. An unsubstantiated method to collection a presumed MST has also been devised here. This type of amalgam dissection procedure which collections the sections its place of twin pixels diminishes importantly the compassion to sound and augments the complete dissection presentation. The dominance of the projected method has been investigated on a great set of multispectral metaphors and associated with some well-known fusion segmentation representations.

Index Terms—Watershed algorithm, image segmentation, mean-shift clustering, minimum spanning tree.

I. INTRODUCTION

Image dissection is one of the rudimentary and primary steps obligatory for great level twin empathetic. It bonds the semantic break between squat level image processing and great level intellectual copy examination. Item gratitude and tracing, land-cover, land-use organization are some of the submissions where separation is one of the essential portions.

Data assembling is one of the entrenched resemblance created approaches which has extensively been functional in the province of twin separation. The main concept of gathering based segmentation is to cluster image landscapes into dissimilar groups so that the intra band disparities are diminished and inter bunch. Numerous journals shot the positive request of many conservative bunching methods (k-means, ambiguous c-means, solidity based collecting) for dissection of satellite imageries. But most of them are parametric and necessitate the estimated original amount of collections to advance auxiliary. Mean-shift gathering system has been discovered in current fiction as a auspicious image segmentation system.

While the assembly founded segmentation methods are effective in discovery original twin landscapes, they execute

some severe problems too. The longitudinal erection and the control statistics of the image are not conserved and pixels from dissimilar provinces are problematic to extricate in instance of meeting article fields. Latitudinal subdivision methods have been discovered in current fiction as a substitute separation approach to sanctuary copy cutoff and latitudinal association among pixels. But the chief difficulty with these procedures is that they unpleasantly yield a big amount of minor quasi-homogeneous areas, i.e. as in the event of crisis alteration. Hence a correct as simulation method is wanted later has future and thing created usual duplicate division organization using super-pixels as a consortium signal.

The super-pixels are attached in a two-part grid separating outline has advanced a fusion division process using crisis alter trailed by a adapted MS built inclusion. The agenda deliberate in has been efficient here in an actual manner. The image is first complete segmented by the future MS based grouping tracked by an integration plan using a new MST based gathering procedure.

MS process is typically practical to realize gap conserving flattening tracked by double dissection. Owing to its control protective sifting stuff the striking landscapes of the general twin are reserved. This stuff is significant for segmenting distantly detected imageries in which numerous different areas are cast-off to signify the entire division. Though, it is tough to divider a distantly detected appearance into dissimilar property shelters merely built on the MS procedure as MS is an unsubstantiated method where the quantity and form of the data groups are indefinite *a priori* however relatively lesser standards of the bandwidth stricture used for the KDE and the minimum section size (M_{min}) allowed can assure over division in the productivity.

Classic grid based division systems are tough to smear straight to image pixels, as the storing and computational charges develop massive. Henceforth, for mixture division algorithms, grid based assembling is the normal excellent for the last area integration resolution any where the bulges of the grid signify the substances removed after the originally ended segmented copy (shaped by some other procedures like MS based division) which are distant fewer in amount than the pixels current in the unique image. Algorithms k-means or Fuzzy are

usually not fortified to be used for this last merger since they incline to yield wired globular groups and they have the intrinsic parametric landscape.

This paper suggests a satellite image division process joining two of the grouping methods stated beyond, i.e., mean-shift and minimum spanning tree based grouping. The innovations of the planned effort are as follows,

- An alteration to the conformist mean-shift clustering has been projected. A k-dist created process for Parzen space breadth approximation system has been established here. For each point, k-dist computes, its reserve to the k th adjacent neighbor. The remoteness quantity at the shrillest evolution fact of the reserve plan is typically measured as the frame size for mass based grouping algorithms. The dismissing standard planned here assurances fast junction of the process.
- A non-parametric grouping system using MST has also been planned. The tree is distributed hierarchically and a distinct gathering is spawned in each repetition. A departure establish has been projected which monitors the grouping course.

The paper is organized as surveys. In Section II, a summary of the projected dissection algorithm is obtainable. The alterations projected for the typical MS built clustering are declared in Unit III. Feature withdrawal from the segmented substances is split. It details the future MST based gathering method. Investigational consequences are deliberated. Finally accomplishes the artifact laterally with the conceivable upcoming postponement of the projected segmentation context.

II. THE PROPOSED IMAGE SEGMENTATION ALGORITHM

Fig. 1 shows the flowchart of the projected segmentation system. Approximately, it is a three-step process.

- Apply the projected improved MS based clustering system to group the pixels of the effort settlement copy in the phantom province.
- A feature abstraction step is approved out to compute some low equal geographies from each of the stuffs found in the former state.
- Since all section as a bulge in the a fresh distinct feature planetary, the anticipated MST created non-parametric gathering method is practical for last expanse integration.

The steps are labeled in detail in succeeding segments.

III. MODIFICATIONS PROPOSED TO THE MEAN-SHIFT TECHNIQUE

The exact particulars of the average mean-shift that should be attained. However rmean -shift fixes not necessitate. The outdated MS grouping process halts when the mass incline

reaches a value near to 0. The projected measure in its places indicates the resident means to produce a universal unchanging flat mass estimator

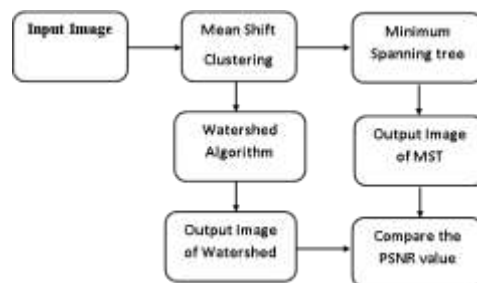


Fig. 1. Flowchart of the proposed segmentation algorithm.

The preliminary quantity of bands essentially current in the statistics cosmos, the breadth of the Parzen space is to be stated previously. If the window extent is superior, there is opportunity of over absorption whereas minor window pointers to the cohort of numerous trivial sub-clusters in the productivity. Here k-dist based process for assessing the (near)- prime bandwidth has been projected. The outmoded mean-shift method stays when the solidity incline reaches a worth near to zero. It is an iterative method and may producer redundant superior amount of groups. Here an adaptive conclusion standard has been advanced which has been verified to quicken the grouping method. This dismissing ailment identifies the sub-clusters existing in a given band first and then a pole dispensation step syndicates them organized into a solitary collection.

Bandwidth Selection using k-Dist

Adjacent Neighbor (AN) founded bandwidth valuation is well branded in the grouping fiction. Here, an allowance of the general k-Dist method (used mostly beside with mass founded. The frame at a period fit to the similar lesson. To certify this, the regular cluster breadth needs to be approached. Hence assumed a point M and x_i , it has been estimated that any of its k th neighbor in the overhead declared series will ultimately fractionous the border of the session where x_i is probable to be extant. The series is empirically fixed and displays good enactment for distantly sensed imageries with average and great longitudinal purposes.

A. Proposed Termination Condition

Preliminary from any arbitrarily designated point x , rendering to the mean-shift perception, the frame changes near the direction of the mostly populated local region in the data space. Beside with the frame size the dissolution disorder is also one of the main issues to be measured to assurance that the procedure does not fall into resident targets.. The suggested procedure is remote extra valuable when a mass is meagerly dispersed in the feature cosmos.

At iteration, contract us deliberate signifies the smidgen of the covariance medium of the facts themes inside the consistent Parzen frame. Let us also study that represents the hint of the covariance medium of the files facts in the frame in the next repetition, i.e. replication.

Quantities sockets fitting to the equal class have less alteration than the facts ideas going to altered lessons. Hence, with the crusade of the frame, once the closure form of attains the smallest assessment, it proposes that the limited type for that band is grasped. So data ideas fitting to all the spaces in this repetition form a single sub-cluster. Now for another iteration, if there is any crossing of points with early labeled points, those two clusters are merged. That is how clusters of any form and scope are found using this adaptive mean-shift grouping. One issue to be lectured here is that, the number of mutual topics in two sequential frames has also been deliberated. The supposition utility certifies that the general situation term resolve achieve the tiniest expected value only at the middle of the frame measure path. These local means (sensed sub-clusters) can be assembled to find a single superior band.

C. Feature Extraction revealed from the Segmented Objects

The substances originate later this phase endure an article abstraction period. Cruel phantom assessment of the section and some numerical quality topographies have been measured for this purpose. Form feature has not been measured here as a given land use class may have dissimilar scope and shape at diverseter restrial position.

Gray level co- existence medium (GLCM) based quality sorts are represented as these kinds of structures can categorize images with mini-textural rudiments. This notion is vital in classifying different land cover classes of remotely sensed satellite imageries. GLCM is defined as the dispersal of the co-occurring ethics within a assumed balance in the image flat. It signifies the reserve and gaunt latitudinal association over an image sub-area of specific size. The GLCM used here is track invariant. Hence the usual of all four latitudinalpreparationsportraying0, 45 has been used. A pixel balance of 1 and a quantization equal of 256 have also been measured. From the co- incidence medium, several surface stricture can be intended. Entropy, dynamism, consistency and disparity have been designated for this case to detention the qualified gray level arbitrariness and intensity of a given section. Therefore for a p -band copy, the areas found after the preceding step are signified in a 2-dimensional feature planetary which are more gathered expending the imminent TST based grouping.

IV. PROPOSED MST BASED CLUSTERING

First construct MST(minimum spanning tree) using the Kruskal algorithm and then set a verge value and step size. then eliminate those boundaries from the MST, whose distances are greater than the threshold value, hence we calculate the ratio among the intra-cluster distance and inter-cluster distance and highest the ratio as well as the threshold.

A. Proposed MST Based Clustering Algorithm

Given a set of points with unidentified underlying probability distribution, a graph is constructed with vertex set and edge set where each of the x_i is represented by a node in the graph. The edge between a given pair of vertices and is weighted by the Euclidean distance has been fixed by grid search technique using 50% of the total data for cross-validation. The value of will be smaller if and are similar to each other in some sense whereas will get increased in case the data points differ.

A minimum spanning tree of is constructed henceforth using the Kruskal's approach .The proposed clustering method checks iteratively whether a given edge of the tree can be deleted to form two compact sub-clusters. The edges are considered in the descending order of the corresponding edge weights because it has been assumed that edges with larger weights are the ones which span different clusters.

It describes the proposed satellite clustering process is a measure of the cluster compactness. If a given edge has drastically larger weight value than the component edge weights of It joining connected components (CC's), then it can be presumed that spans two different clusters and hence can be deleted. The parameter of defined tolerance per CC. Given a CC, the corresponding α controls the gap allowed in the feature space between this CC and any other CC.

B. Efficiency Assessment of the MST Based Clustering

The goodness of the projected clustering has been restrained in terms of the Davies-Putnam (DP) index _{α} and Silhouette (Sil) index. The value of DB index is lowest at the finest clustering point whereas Sil index attains the maximum value at this point. The value of these two catalogues for the proposed clustering have been compared with those of some of the well-known clustering (parametric and non-parametric) techniques.

V. EXPERIMENTAL RESULTS

Four study areas have been presented here for analysis. The images considered here are mainly of the Karnataka Area, India acquired by Indian Isolated Detecting Satellite Linear

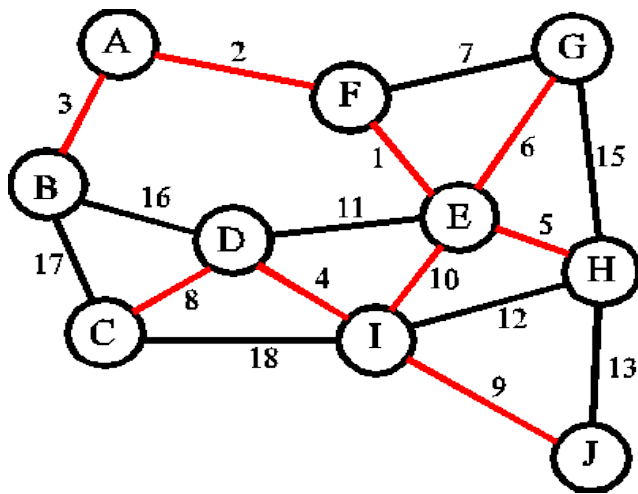


Fig 2. Minimum spanning tree feature extraction

Validity Assessment of the Proposed MST Clustering

The objects extracted after the proposed MS based segmentation undergo a feature extraction stage. The objects are further merged using the proposed MST based clustering technique. The efficiency of this clustering has been compared with some well-known clustering techniques from the literature, i.e., normalized cut based graph clustering [20], adaptive mean-shift based clustering with data centric bandwidth estimation, means++ [21] and Fuzzy *c*-means in terms of Silhouette index and DB index. The initial cluster centers selected by means++ have been used for Fuzzy *c*-means also which is based upon the intuition of spreading the initial clusters away from each other. Tables IV and V show the results of the comparison based on cluster validity indices. For normalized cut based clustering, means++ and Fuzzy, *c*-means, the number of clusters (N_c) has been set physically to the rough number of classes present in the image based on expert opinion. The values mentioned here are the average of 25 realizations.

N_{MST} Symbolizes the number of clusters produced by the proposed MST based clustering. It can be observed from both Tables IV and V that the proposed clustering technique performs near-optimal clustering in almost all the cases and is comparable to other popular clustering techniques. The problem with techniques like Normalized cut, means++ or Fuzzy *c*-means is that the approximate number of clusters needs to be supplied beforehand and even under-merging occurs with those methods. It is difficult to ascertain this information from remotely sensed satellite images *a priori*. The N_{MST} measures exhibit that the proposed method can

detect the actual number of clusters. Problems due to cluster shape and size are well taken care of by the underlying tree structure.

E. Comparison of the Segmentation Results

The segmentation results of the proposed hybrid clustering based method have been compared with two recent object

TABLE VI
COMPARISON IN TERMS OF OVERALL ACCURACY (IN %) (FOR IMAGES IN FIG. 2(a)–2(d))

Proposed method	Watershed + MS based	MS + Normalized cut based
85.34	73.68	77.24
84.62	81.11	82.33
79.41	76.29	80.70
81.67	79.32	76.61

Based clustering techniques [10] and [8]. Like [10], [8] is also a two-step process where the objects extracted from the over-segmented output of the MS based clustering are merged using Normalized cut based clustering. The accuracy assessment for the segmentation technique is performed by comparing the results to the reference classified images obtained by expert annotation.

From the segmentation output of Fig. 2(i), it is evident that the proposed segmentation is able to detect the underlying land-cover classes with maximum possible merging. The outputs in Fig. 2(m) and (q) are largely over-segmented and a certain amount of mis-classification can be observed in both cases. The proposed method also preserves the local texture properties (Fig. 2(j) and (l)). The corresponding results obtained by the other algorithms (Fig. 2(n), (p), (r) and (t)) are over-merged to some extent, hence the local texture details are difficult to ascertain.

These three segmentation methods have further been compared with respect to the reference manually classified image in terms of overall classification accuracy. The reference images have been labeled manually using the domain knowledge. In a classification matrix of size where denotes the number of clusters, a typical entry defines how many samples of class *i* have been misclassified to class *j*. The overall classification accuracy is defined as (5) N is the number of samples in the data set. Table VI shows the accuracy assessment of the proposed segmentation technique with respect to the other ones used in terms of $\zeta_{overall}$.

Overall, the proposed segmentation algorithm is capable of segmenting a given satellite image into almost accurate number of land cover classes without any user intervention. Use of MS algorithm with properly tuned parameters guarantees an over segmented image while keeping fine image details. The edge preserving smoothing property of MS based segmentation helps in removing unnecessary noise elements

that may be present in the image due to improper sensing or environmental hazards.

Hence, MS based clustering is a better option than the traditional watershed based segmentation in the initial stage. Now to merge these initially found objects, it is always better to adopt any pure non-parametric method due to lack of domain knowledge available for a given remotely sensed image. Non-parametric algorithms like MS are not encouraged to use for this merging stage as these algorithms are often stuck in local modes present in the feature space. The proposed MST based clustering does not suffer from such kind of problems and furthermore, the proposed separation predicate allows sub-division up to a level when the clusters under consideration are sufficiently far from each other.

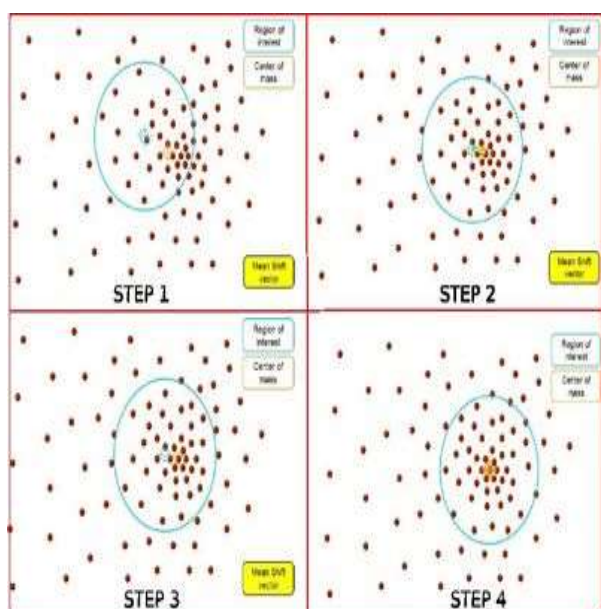


Fig 3: mean shift clustering segmentation on particles

Mean Shift is a powerful and adaptable non parametric iterative algorithm that can be used for lot of purposes like discovery modes, clustering etc. Fix a window around each data point. Calculate the mean of data within the window. Shift the window to the mean and repeat till meeting.

This MST based clustering does not need the number of clusters to be stated a priori. A novel dismissing criteria and a original method for guessing the Parzen window width have been extra to the traditional MS clustering algorithm which assurances fast convergence while accomplishment near-optimal bunching by the MS algorithm. Parametric and require the estimated initial number of bunches to ensue further. The spatial structure and the edge material of the image are not preserved and pixels from different regions are problematic to distinguish in case of overlapping feature domains. Unpleasantly produce a large number of minor quasi-homogeneous regions.

VI. CONCLUSION

An original unconfirmed settlement duplicate subdivision algorithm has been proposed here. It consists of two parts. First the image is segmented in the ethereal province by a improved mean-shift based clustering technique. A novel terminating measures and a original system for estimating the Parzen window width have been added to the outdated MS clustering algorithm which assurances fast conjunction while performing near-optimal clustering by the MS algorithm. The output of this step is an over segmented version of the original image. A feature quarrying is performed to extract color and surface features from each object thus extracted. The objects are clustered in the feature space using a non-parametric minimum straddling tree based clustering method. This MST based clustering does not require the number of clusters to be mentioned *a priori*. The algorithm preserves the view of object based image analysis and is comparable to other object based segmentation techniques. The experiments performed show that, the method exhibits healthier classification accuracy than methods like Watershed + MS based segmentation or MS + Standardized cut based segmentation methods. This segmentation can be used further to extract specific objects like roads, constructions from the image. Object based image understanding which requires clean and well-organized abstraction of image regions is additional domain where the proposed method can augment exactness to some extent. To segment satellite images with high accuracy using methods that shows better classification accuracy.

REFERENCES

- [1] Q. Hao, R. Cai, Z. Li, L. Zhang, Y. Pang, and F. Wu, "3D visual phrases for landmark recognition," in *Proc. 2012 IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3594–3601.
- [2] W. Long, III and S. Srihann, "Land cover classification of SSC image: Unsupervised and supervised classification using Erdas Imagine," in *Proc. 2004 IEEE Geoscience and Remote Sensing Symp.*, 2004, vol. 4, pp. 2707–2712.
- [3] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Computing Surveys (CSUR)*, vol. 31, no. 3, pp. 264–323, 1999.
- [4] J.-F. Yang, S.-S. Hao, and P.-C. Chung, "Color image segmentation using fuzzy C-means and eigenspace projections," *Signal Process.*, vol. 82, no. 3, pp. 461–472, 2002.
- [5] S.Saha and S.Bandyopadhyay, "Application of a new symmetry-based cluster validity index for satellite image segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 166–170, 2008.
- [6] C. Wemmert, A. Puissant, G. Forestier, and P. Gancarski, "Multiresolution remote sensing image clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 3, pp. 533–537, 2009.
- [7] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [8] W. Tao, H. Jin, and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts," *IEEE Trans. Syst., Man, Cybern., B: Cybernetics*, vol. 37, no. 5, pp. 1382–1389, 2007.
- [9] Z. Li, X.-M. Wu, and S.-F. Chang, "Segmentation using superpixels: A bipartite graph partitioning approach," in *Proc.*

- 2012 IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2012, pp. 789–796.
- [10] B. Banerjee, V. Surender, and K. Buddhiraju, “Satellite image segmentation: A novel adaptive mean-shift clustering based approach,” in *Proc. 2012 IEEE Geoscience and Remote Sensing Symp. (IGARSS)*, Jul. 2012, pp. 4319–4322.
- [11] B. Georgescu, I. Shim shoni, and P. Meer, “Mean shift based clustering in high dimensions: A texture classification example,” in *Proc. 9th IEEE Int. Conf. Computer Vision*, 2003, pp. 456–463.
- [12] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proc. 2nd Int. Conf. Knowledge Discovery and Data Mining*, 1996, vol. 1996, pp. 226–231, AAAI Press.
- [13] R. Panigrahy, “An improved algorithm finding nearest neighbor using kd-trees,” *LATIN 2008: Theoretical Informatics*, pp. 387–398, 2008.
- [14] R. M. Haralick, “Statistical and structural approaches to texture,” *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, 1979.
- [15] O. Grygorash, Y. Zhou, and Z. Jorgensen, “Minimum spanning tree based clustering algorithms,” in *Proc. 18th IEEE Int. Conf. Tools With Artificial Intelligence, ICTAI'06*, 2006, pp. 73–81.
- [16] C.-W. Hsu, C.-C. Chang, and C.-J. Lin *et al.*, “A Practical Guide to Support Vector Classification,” 2003 [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- [17] J. B. Kruskal, “On the shortest spanning subtree of a graph and the traveling salesman problem,” *Proc. American Mathematical Society*, vol. 7, no. 1, pp. 48–50, 1956.
- [18] C. Xia, W. Hsu, M. L. Lee, and B. C. Ooi, “Border: Efficient computation of boundary points,” *IEEE Trans. Knowledge Data Eng.*, vol. 18, no. 3, pp. 289–303, 2006.
- [19] D. Comaniciu, “An algorithm for data-driven bandwidth selection,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 281–288, 2003.
- [20] J. Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [21] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding,” in *Proc. 18th Annual ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.