IMPACT: International Journal of Research in Engineering & Technology (IMPACT: IJRET)
ISSN 2321-8843
Vol. 1, Issue 3, Aug 2013, 109-120

SURVEY PAPER ON VARIOUS METHODS IN CONTENT BASED INFORMATION RETRIEVAL

S. SUBITHA[1] & S. SUJATHA[2]

[1]Department of Computer Applications, M. A. M College of Engineering, Siruganur, Trichy, Tamil Nadu, India

[2]Department of Computer Applications, University Department of Anna University, Tiruchirappalli BIT Campus, Tiruchirappalli, Tamil Nadu, India

ABSTRACT

Information Retrieval is an emerging research area in the field of Information Retrieval. Due to the immense amount of data in the WWW, it is very tough for the user to retrieve the relevant images. Traditional Image Retrieval approaches based on topic similarity alone is not sufficient nowadays the content based image retrieval (CBIR) are becoming a source of exact and fast retrieval. A variety of techniques have been developed to improve the performance of CBIR. Data clustering is an unsupervised method for extraction hidden pattern from huge data sets. With large data sets, there is possibility of high dimensionality. Having both accuracy and efficiency for high dimensional data sets with enormous number of samples is a challenging arena. In this paper the clustering techniques are discussed and analysed. Also, we propose a method HDK that uses more than one clustering technique to improve the performance of CBIR.This method makes use of hierachical and divide and conquer K- Means clustering technique with equivalency and compatible relation concepts to improve the performance of the K-Means for using in high dimensional datasets. It also introduced the feature like color, texture and shape for accurate and effective retrieval system. This survey gives an introduction to content-based image Retrieval and explores the different types of retrieval methods

KEYWORDS: CBIR, Image Feature Extraction, Image Analysis, Image Retrieval, Image Similarity Clustering Techniques

INTRODUCTION

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases (see this survey[1] for a recent scientific overview of the CBIR field). Content-based image retrieval is opposed to concept-based approaches .

"Content-based" means that the search analyzes the contents of the image rather than themetadata such as keywords, tags, or descriptions associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because most web-based image search engines rely purely on metadata and this produces a lot of garbage in the results Also having humans manually enter keywords for images in a large database can be inefficient, expensive and may not capture every keyword that describes the image. Thus a system that can filter images based on their content would provide better indexing and return more accurate results.

The term "content-based image retrieval" seems to have originated in 1992 when it was used by T. Kato to describe experiments into automatic retrieval of images from a database, based on the colors and shapes present.[2] Since then, the term has been used to describe the process of retrieving desired images from a large collection on the basis of syntactical image features. The techniques, tools, and algorithms that are used originate from fields such as statistics, pattern recognition, signal processing, and computer vision.

There is a growing interest in CBIR because of the limitations inherent in metadata-based systems, as well as the large range of possible uses for efficient image retrieval. Textual information about images can be easily searched using existing technology, but this requires humans to manually describe each image in the database. This is impractical for very large databases or for images that are generated automatically, e.g. those from surveillance cameras. It is also possible to miss images that use different synonyms in their descriptions. Systems based on categorizing images in semantic classes like "cat" as a subclass of "animal" avoid this problem but still face the same scaling issues.

CBIR TECHNIQUES

Many CBIR systems have been developed, but the problem of retrieving images on the basis of their pixel content remains largely unsolved.

Query Techniques

Different implementations of CBIR make use of different types of user queries. Query by example is a query technique that involves providing the CBIR system with an example image that it will then base its search upon. The underlying search algorithms may vary depending on the application, but result images should all share common elements with the provided example.

Options for providing example images to the system include:

- A preexisting image may be supplied by the user or chosen from a random set.
- The user draws a rough approximation of the image they are looking for, for example with blobs of color or general shapes.

This query technique removes the difficulties that can arise when trying to describe images with words.

Semantic Retrieval

The ideal CBIR system from a user perspective would involve what is referred to as *semantic* retrieval, where the user makes a request like "find pictures of Abraham Lincoln". This type of open-ended task is very difficult for computers to perform - pictures of chihuahuas and Great Danes look very different, and Lincoln may not always be facing the camera or in the same pose. Current CBIR systems therefore generally make use of lower-level features like texture, color, and shape, although some systems take advantage of very common higher-level features like faces . Not every CBIR system is generic. Some systems are designed for a specific domain, e.g. shape matching can be used for finding parts inside a CAD-CAM database.

Other Query Methods

Other query methods include browsing for example images, navigating customized/hierarchical categories, querying by image region (rather than the entire image), querying by multiple example images, querying by visual sketch, querying by direct specification of image features, and multimodal queries (e.g. combining touch, voice, etc.)

CBIR systems can also make use of *relevance feedback*, where the user progressively refines the search results by marking images in the results as "relevant", "not relevant", or "neutral" to the search query, then repeating the search with the new information.

Content Comparison Using Image Distance Measures

The most common method for comparing two images in content-based image retrieval (typically an example image and an image from the database) is using an image distance measure. An image distance measure compares the similarity of two images in various dimensions such as color, texture, shape, and others. For example a distance of 0 signifies an exact match with the query, with respect to the dimensions that were considered. As one may intuitively gather, a value greater than 0 indicates various degrees of similarities between the images. Search results then can be sorted based on their distance to the queried image.[3] A long list of distance measures can be found in.

Color

Computing distance measures based on color similarity is achieved by computing a color histogram for each image that identifies the proportion of pixels within an image holding specific values (that humans express as colors). Current research is attempting to segment color proportion by region and by spatial relationship among several color regions. Examining images based on the colors they contain is one of the most widely used techniques because it does not depend on image size or orientation. Color searches will usually involve comparing color histograms, though this is not the only technique in practice.

Texture

Texture measures look for visual patterns in images and how they are spatially defined. Textures are represented by texels which are then placed into a number of sets, depending on how many textures are detected in the image. These sets not only define the texture, but also where in the image the texture is located.

Texture is a difficult concept to represent. The identification of specific textures in an image is achieved primarily by modeling texture as a two-dimensional gray level variation. The relative brightness of pairs of pixels is computed such that degree of contrast, regularity, coarseness and directionality may be estimated (Tamura, Mori & Yamawaki, 1978). However, the problem is in identifying patterns of co-pixel variation and associating them with particular classes of textures such as silky, or rough.

Shape

Shape does not refer to the shape of an image but to the shape of a particular region that is being sought out. Shapes will often be determined first applying segmentation or edge detection to an image. Other methods like [Tushabe and Wilkinson 2008] use shape filters to identify given shapes of an image. In some case accurate shape detection will require human intervention because methods like segmentation are very difficult to completely automate.

THE RETRIEVAL BASED ON CLUSTERING TECHNIQUES

Clustering techniques can be classified into supervised (including semi-supervised) and unsupervised schemes. The former consists of hierarchical approaches that demand human interaction to generate splitting criteria for clustering. In unsupervised classification, called clustering or exploratory data analysis, no labeled data are available The goal of clustering is to separate a finite unlabeled data set into a finite and discrete set of “natural,” hidden data structures, rather than provide an accurate characterization of unobserved samples generated from the same probability distribution This paper critically reviews and summarizes different clustering techniques.

Log –Based Clustering

Images can be clustered based on the retrieval system logs maintained by an information retrieval process. The session keys are created and accessed for retrieval. Through this the session clusters are created. Each session cluster generates log –based document and similarity of image couple is retrieved. Log –based vector is created for each session vector based on the log-based document. Now, the session cluster is replaced with this vector. The unaccessed documents creates its own vector.

A hybrid matrix is generated with at least one individual document vector and one log-based clustered vector. At last the hybrid matrix is clustered. This technique is difficult to perform in the case of multidimensional images. To overcome this hierarchical clustering is adopted.

Hierarchical Clustering

Hierarchical clustering (HC) algorithms organize data into a hierarchical structure according to the proximity matrix. The results of HC are usually depicted by a binary tree or dendrogram as shown in Figure 1where A, B, C, D, E, F, G are objects or clusters. It represents the nested grouping of patterns and similarity levels at which groupings change. The root node of the dendrogram represents the whole data set and each leaf node is regarded as a data object. The intermediate nodes, thus, describe the extent that the objects are proximal to each other; and the height of the dendrogram usually expresses the distance between each pair of objects or clusters, or an object and a cluster. The ultimate clustering results can be obtained by cutting the dendrogram at different levels. This representation provides very informative descriptions and visualization for the potential data clustering structures, especially when real hierarchical relations exist in the data, like the data from evolutionary research on different species of organizms. HC algorithms are mainly classified as agglomerative methods and divisive methods. Agglomerative clustering starts with clusters and each of them includes exactly one object. A series of merge operations are then followed out that finally lead all objects to the same group. Divisive clustering proceeds in an opposite way. In the beginning, the entire data set belongs to a cluster and a procedure successively divides it until all clusters are singleton clusters. For a cluster with objects, there are 2^{N-1}-1 possible two-subset divisions, which is very expensive in computation. Therefore, divisive clustering is not commonly used in practice. In recent years, with the requirement for handling large-scale data sets in data mining and other fields, many new HC techniques have appeared and greatly improved the clustering performance.

Retrieval Dictionary Based Clustering

A rough classification retrieval system is formed. This is formed by calculating the distance between two learned patterns and these learned patterns are classified into different clusters followed by a retrieval stage. The main drawback addressed in this system is the determination of the distance. To overcome this problem a retrieval system is developed by retrieval dictionary based clustering. This method has a retrieval dictionary generation unit that classifies learned patterns into plural clusters and creates a retrieval dictionary using the clusters. Here, the image is retrieved based on the distance between two spheres with different radii. Each radius is a similarity measure between central cluster and an input image. An image which is similar to the query image will be retrieved using retrieval dictionary.

NCut Algorithm

Ncut method attempts to organize nodes into groups so that the within the group similarity is high, and/or between the groups similarity is low. This method is empirically shown to be relatively robust in image segmentation. This method can be recursively applied to get more than two clusters. In this method each time the subgraph with maximum number of nodes is partitioned (random selection for tie breaking). The process terminates when the bound on the number of clusters is reached or the Ncut value exceeds some threshold T. The recursive Ncut partition is essentially a hierarchical divisive clustering process that produces a tree. Nonetheless, the tree organization here may misleading a user because there is no guarantee of any correspondence between the tree and the semantic structure of images. Furthermore, organizing image clusters into a tree structure will significantly complicate the user interface.

K Means Clustering

This nonhieraircha1 method initially takes the number of components of the population equal to the final required number of clusters. In this step itself the final required number of clusters is chosen such that the points are mutually farthest apart. Next, it examines each component in the population and assigns it to one of the clusters depending on the minimum distance. The centroid's position is recalculated everytime a component is added to the cluster and this continues until all the components are grouped into the final required number of clusters.The K- means algorithm is very simple and can be easily implemented in solving many practical problems. It can work very well for compact and hyperspherical clusters. The time complexity of K-means is O (NKd). Since K and d are usually much less than N,K-means can be used to cluster large data sets. Parallel techniques for K-means are developed that can largely accelerate the algorithm. Incremental clustering techniques for example (Bradley et al., 1998) do not require the storage of the entire data set, and can handle it in a one-pattern-at-a-time way. If the pattern displays enough closeness to a cluster according to some predefined criteria, it is assigned to the cluster. Otherwise, a new cluster is created to represent the object.

Graph Theory Based Clustering

The concepts and properties of graph theory make it very convenient to describe clustering problems by means of graphs. Nodes of a weighted graph correspond to data points in the pattern space and edges reflect the proximities between each pair of data points. A graph-based clustering method is particularly well suited for dealing with data that is used in the construction of minimum spanning tree MST. It can be used for detecting clusters of any size and shape without specifying the actual number of clusters. Well known algorithms in clustering are Minimum Spanning Tree based clustering, and clustering editing method, HCS algorithm, etc. Current research is focused on clustering using divide and conquers approach. Usually this clustering methodology is used to detect irregular clustering boundaries in clustering results. Zhan proposes to construct an MST and delete the inconsistent edges, i.e. the edges weight values are significantly larger than average weight of the nearby edges in the tree. The inconsistency measure is applied to each edge to detect and remove the inconsistence edges, which results as a set of disjoint subtrees, each subtree will represent a separate cluster

Divide and Conquer K-Means

When the size of a data set is too large, it is possible to divide the data into different subsets and to use the selected cluster algorithm separately to these subsets. This approach is known as divide and conquer . The divide and conquer algorithm first divides the entire data set into a subset based on some criteria. The selected subset is again clustered with a clustering algorithm K-Means. The advantage is to accelerate search and to reduce complexity which depends on number of samples. Methods based on *subspace clustering* may help to ease the problem of clustering high-dimensional data, but they are not adapted at obtaining a large number of clusters .A possible solution to this issue, is to cluster hierarchically (obtain a small number of clusters and then cluster again each of the clusters obtained). The proposed enhanced clustering method HDK which uses the combination of unsupervised clustering methods is one of the method that can largely accelerate the CBIR system.

CONCLUSIONS

The purpose of this survey is to provide an overview of the functionality of content based image retrieval systems. Combining advantages of HC and divide and conquer K-Means strategy can help us in both efficiency and quality. HC algorithm can construct structured clusters. Although HC yields high quality clusters but its complexity is quadratic and is not suitable for huge datasets and high dimension data. In contrast K-Means is linear with size of data set and dimension and can be used for big datasets that yields low quality. Divide and conquer K-Means can be used for high dimensional data set . In this paper we present a method HDK to use both advantages of HC and Divide and conquer K-Means by introducing equivalency and compatible relation concepts. Using two steps clustering in high dimensional data sets with considering no of clusters based on color feature helps us to improve accuracy and efficiency of original K-Means clustering. For this purpose we should consider orthogonal space. HDK algorithm has been used extensively in various areas to improve the performance of the system and to achieve better results in different applications.

REFERENCES

1. Rui Xu, Donald Wunsch II, “Survey of Clustering Algorithms”, IEEE Transactions On Neural Networks, Vol. 16, No. 3, May 2005.
2. Ritendra Datta, Dhiraj Joshi, Jia Li, And James Z. Wang, “Image Retrieval: Ideas, Influences, and Trends of the New Age”, ACM Computing Surveys, Vol. 40, No. 2, Article 5, Publication date: April 2008.
3. Tamura et al. “Texture Feautres Corresponding to Visual Perception”- IEEE Trans on system, Man and cyber 8-460-472-1978.
4. Sanjoy Kumar Saha et al. “CBIR Using Perception Based Texture And Colour measures ”CSE Department; CST Department Jadavpur Univ., India; B.E. College, Unit ISI, Kolkata, India - -2003.
5. Canny, J., “A computational approach to edge detection”, IEEE Trans on Pattern Analysis and Machine Intelligence, 8:679-698, 1986.
6. S. Nandagopalan, Dr. B.S. Adiga, and N. Deepak “A Universal Model for Content-Based Image Retrieval” World Academy of Science, Engineering and Technology 46 2008.
7. Mohammed Eisa and Ibrahim Elhenawy and A. E. Elalfi and Hans Burkhardt, “Image Retrieval based on Invariant Features and Histogram Refinement”, ICGST International Journal on Graphics, Vision and Image Processing, March 2006, pp. 7-11.
8. V. Castelli and L. D. Bergman (Eds.), “Image Databases: Search and Retrieval of digital Imagery”, Wiley: New York, 2002.
9. B. Everitt, S. Landau, and M. Leese, “Cluster Analysis”, London:Arnold, 2001.
10. Jain and R. Dubes, “Algorithms for Clustering Data”, Englewood Cliffs, NJ: Prentice-Hall, 1988.
11. A. Baraldi and E. Alpaydin, “Constructive feedforward ART clustering networks—Part I and II,” IEEE Trans. Neural Netw., vol. 13, no. 3, pp.645–677, May 2002.
12. V. Cherkassky and F. Mulier, “Learning From Data: Concepts, Theory, and Methods”, New York: Wiley, 1998.

13. I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas and P.Yianilos, “The Bayesian Image Retrieval System, PicHunter: Theory, Implementation and Psychophysical Experiments”, IEEE Trans. Image Processing, vol. 9, no. 1, pp. 20-37, 2000.
14. Y. Rui, T. S. Huang, M. Ortega and S. Mehrotra, “Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval”, IEEE Trans. Circuits and Video Technology, vol. 8, no. 5, pp. 644-655, 1998.
15. S. Thilagamani and N. Shanthi **, “**A Survey on Image Segmentation Through Clustering”, International Journal of Research and Reviews in Information Sciences , Vol. 1, No. 1, March 2011 .
16. Rui Y, Thomas, S. Huang. “Content-based image retrieval with relevance feedback in MARS”, In proceedings of IEEE international conference on image processing, pp. 815-818, 1997.
17. Hoi, C.-H. and Lyu, M. R. 2004b. “A novel log based relevance feedback technique in content based image retrieval”. In Proc. ACM Multimedia,2004.
18. Zhou XS, Huang TS., Relevance feedback in image retrieval: A comprehensive review. Multimedia Syst; 8: 536-544, 2003.
19. P.Chundi, U.Dayal, , M.Sayal,, M.Hsu, “*A Document Clustering Method and System*:” google patents,US20077181678, 2007.
20. D. Melas and S. Wilson, “Double Markov Random Fields and Bayesian Image Segmentation,” IEEE Trans. Signal Processing,vol. 50, no. 2, pp. 357-365, Feb. 2002.
21. Huiyu Zhou, Abdul H. Sadka, Mohammad R. Swash, Jawid Azizi and Abubakar S. Umar., “Content Based Image Retrieval and clustering: A Brief Survey” school of Engineering and Design, Brunel University, Uxbridge, UB8 3PH, UK
22. Hoi, C.-H. and Lyu, M. R. 2004a. “Group-based relevance feeedbacks with support vector machine ensembles” In Proc. IEEE ICPR,2004.
23. Li Wenchao Zhou Yong Xia Shixiong China Univ. of Min. & Technol., Xuzhou, “A Novel Clustering Algorithm Based on Hierarchical and K-means Clustering” Print ISBN: 978-7-81124-055-9, On page(s): 605, 2009.
24. Jianbo Shi Malik, J.Robotics Inst., Carnegie Mellon Univ., Pittsburgh, PA , “Normalized cuts and image segmentation”, ISSN: 0162-8828, On page(s): 888 – 905,2000.
25. Emma Regentova, Dongsheng Yao, and Shahram Latifi, “Image segmentation using Ncut in the wavelet domain”, International journal of image and graphics,world scientific Publishing company,vol.6(4),pp.569-582,2006.
26. Guha, Meyerson, A. Mishra, N. Motwani, and O. C. ."Clustering data streams: Theory and practice ." *IEEE* Transactions on Knowledge and Data Engineering, vol. 15,pp. 515-528, 2003.
27. A. Jain, M. Murty, and p. Flynn " Data clustering: A review.," ACM Computing Surveys, vol. 31, pp. 264-323,1999.
28. L. Parsons, E. Haque and H. Liu., “Subspace clustering for high dimensional data: a review.” *SIGKDD Explor. Newslett*. 6 (1), pp. 90–105. 2004.

29. C. Bouveyron, S. Girard and C. Schmid, “High-dimensional data clustering. Computational Statistics & Data Analysis”, vol. 52, pp.502–519. 2007.

30. B. Everitt, S. Landau, and M. Leese, “Cluster Analysis” . London:Arnold, 2001., [31]

31. Classification, 2nd ed. London, U.K.: Chapman & Hall, 1999.

32. P. Hansen and B. Jaumard, “Cluster analysis and mathematical programming”, Math. Program., vol. 79, pp. 191–215, 1997.

33. A. Jain and R. Dubes, “Algorithms for Clustering Data”. Englewood Cliffs, NJ: Prentice-Hall, 1988. [34]

34. Greg Pass, Ramin Zabih, “Histogram refinement for content based image retrieval” WACV '96.

35. Chad Carson, Serge Belongie, Hay it Greenspan, and Jitendra Malik, “Region Based Image Querying,” this work is supported by an NSF digital library grant (IRI 94-11334) 1997 IEEE.

36. Minakshi Banerjee 1, Malay K. Kundu “Edge based features for content based image retrieval” 0031-3203/2003.

37. M. J. Swain and D. H. Ballard, “Color indexing” , *Int. d. Comput. Vis.* 7(1), 11-32 (1991).

38. Mohan S. Kankanhalli,T Babu M. Mehtre and Jian Kang Wut, “Cluster-Based Color Matching”, Pattern Recognition. Vol. 29, No. 4, pp. 701 708, 1996.

39. Madjid Khalilian, Norwati Mustapha, MD Nasir Suliman, MD Ali Mamat,” A Novel K-Means Based Clustering Algorithm for High Dimensional Data Sets”,Proceedings of the IMECS 2010.

40. Madjid Khalilian, Farsad Zamani Boroujeni, Norwati Mustapha, Md. Nasir Sulaiman, “K-Means Divide and Conquer Clustering”, International Conference on Computer and Automation Engineering.

41. H. T. Shen, X. Zhou, and A. Zhou. “An adaptive and dynamic dimensionality reduction method for high-dimensional indexing” VLDB J., 16(2):219{234,2007.

42. Y. Tao, K. Yi, C. Sheng, and P. Kalnis., “Quality and efficiency in high dimensional nearest neighbor search”,.In SIGMOD, pages 563{576, 2009.

43. R. Weber, H.-J. Schek, and S. Blott., “A quantitative analysis and performance study for similarity- search methods in high-dimensional spaces.”, In VLDB, pages 194{205, 1998.

44. C. Bouveyran, S. Girard , C.Schmid : Technical Report 1083M, LMC-IMAGE 2008.

45. C. Fraley and A. Raftery, “Model based clustering discriminate analysis and density estimation”, journal of American statistics association, 97, 611-631 ,2002.

46. H. Bock, on the interface between cluster analysis, principal component clustering and multidimensional scaling, 1987.

47. Jeffery L. Solka: Text data mining Theory and methods, vol 2, 94-112, 2008

48. H. Huang, C. Ding, D. Luo, T. Li, “simultaneous tensor subspace selection and clustering”, KDD08, lasvegas USA ,2008.

49. J. Ye, Z. Zhao, M. Wu, “Discriminative K-Means for clustering”, proceeding of the annual conference, 2007.

50. R. Varshavsky, D. itorn and M. Linial, "Cluster algorithm optimizer: A framework for large datasets" ,ISBRA pp 85-96,springer 2007.
51. W. Tang, H. Xiong, S. Zhang, J. Wu, "Enhancing semisupervised clustering", KDD07 california USA , ACM 2007.
52. Yuntao Qian and Ching Y. Suen , "Clustering Combination Method", IEEE 2000.
53. Yuntao Qian and Ching Y. Suen, "Sequential combination methods for data clustering", Journal of computer science,2002.
54. A. Raftery and N. Dean, "variable selection for model based clustering", journal of American statistical association, 101(473):168-178,2006.
55. Pankaj K. Agarwal and Nabil H. Mustafa, "K-Means Projective clustering", PODS Paris France, ACM 2004.
56. L. Fred and A. K. Jain., "Combining multiple clusterings using evidence accumulation", IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(6):835–850, 2005.
57. Ivan O.Kyrgyzov, Henri Maˆıtre and Marine Campedel, "A Method of Clustering Combination Applied to Satellite Image Analysis".
58. A. P. Topchy, A. K. Jain, and W. F. Punch., "A mixture model for clustering ensembles", In *SDM*, 2004.
59. T. Lange and J.M. Buhmann, "Combining partitions by probabilistic label aggregation", In KDD '05: Proceeding of theeleventh ACM SIGKDD international conference on Knowledge discovery in data mining, pages 147–156, New York,NY, USA, 2005. ACM Press.
60. Christian Hartvedt, "Using Context to Understand User Intentions in Image Retrieval" 2010.
61. ZhiYong Zeng, ZhiQiang Yao, ShiGang Liu, "An Efficient and effective Image Representation for Region-Based Image Retrieval" ICIS 2009, November.
62. Qifeng Qiao, Peter A. Beling, "Localized Content Based Image Retrieval with Self-taught Multiple Instance Learning" 2009 IEEE.
63. Nidhi Singhai, Prof. Shishir K. Shandilya, "A Survey On: Content Based Image Retrieval Systems", International Journal of Computer Applications (0975 – 887)Volume 4 – No.2, July 2010.
64. Steven M. Holand , "Cluster Analysis", Department of Geology, University of Georgia, Athens, GA 30602-2501
65. S. Guha, R. Rastogi, and K. Shim, "CURE: An efficient clustering algorithm for large databases," in*Proc. ACM SIGMOD Int. Conf. Management of Data*, 1998, pp. 73–84.
66. "ROCK: A robust clustering algorithm for categorical attributes," *Inf. Syst.*, vol. 25, no. 5, pp. 345–366, 2000.
67. G. Karypis, E. Han, and V. Kumar, "Chameleon: Hierarchical clustering using dynamic modeling," *IEEE Computer*, vol. 32, no. 8, pp. 68–75,Aug. 1999.
68. T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An efficient data clustering method for very large databases," in *Proc. ACM SIGMOD Conf. Management of Data*, 1996, pp. 103–114.

69. Ball, G., Hall, D., 1965. ISODATA, a novel method of data anlysis and pattern classification. Technical report NTIS AD 699616. Stanford Research Institute, Stanford, CA.

70. E. Dahlhaus, "Parallel algorithms for hierarchical clustering and applications to split decomposition and parity graph recognition," *J. Algorithms*, vol. 36, no. 2, pp. 205–240, 2000.

71. C. Olson, "Parallel algorithms for hierarchical clustering," *Parallel Comput.*, vol. 21, pp. 1313–1325, 1995.

72. K. Stoffel and A. Belkoniene, "Parallel K-means clustering for large data sets," in *Proc. EuroPar'99 Parallel Processing*, 1999, pp.1451–1454.

73. F. Harary, *Graph Theory*. Reading, MA: Addison-Wesley, 1969

74. C.T Zahn(1971) : Graph-theoretical methods for detecting and describing clusters , IEEE trans on Computers c(20):68-86

75. G. Karypis; E.Han(1999) : "A hierarchical clustering Algorithm using dynamic modeling, IEEE Trans on Computers:", Special issue on Data analysis and Mining, 32(8):68-75

76. J.Gramm; J.Guo(2003) : Graph modeled data clustering: fixed parameter algorithms for clique generation, In Lecture Noted on Computer Science(LNCS),pages 109-118,Springer

77. R. R.Shamir; D.Tsur(2002): Cluster graph modification problems. In Lecturer notes in computer science(LNCSI),pp379-390,springer

78. E. Hartuy ; R. Sharmir. A clustering algorithm based on graph connectivity. Information processing,pp175-181

79. Xiaochun Wang; D. Mitchell Wilkes (2009) : "A Divide-and-Conquer Approach for Minimum Spanning tree-Based Clustering", Member,IEEE Transactions on knowledge and data engineering, Vol 21 No7

80. Oleksandr; Grygorash(2006); Yan Zhou: "Minimum Spanning tree Based Clustering", IEEE Tools with artificial intelligence,pp 3-81

81. Guttman, R-tree: "A dynamic index structure for spatial searching", in *Proc. ACM SIGMOD, 1984*.

82. T. Sellis, N. Roussopoulos, and C. Faloutsos, "The RC-tree: A dynamic index for multi-dimensional objects", in *Proc. 12th VLDB, 1987*.

83. D. Greene, "An implementation and performance analysis of spatial data access", in *Proc. ACM SIGMOD,1989.*

84. N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger, The R¤-tree: "An efficient and robust access method for points and rectangles", in *Proc. ACM SIGMOD, 1990*.

85. Y. Rui, K. Chakrabarti, S. Mehrotra, Y. Zhao, and T. S. Huang, "Dynamic clustering for optimal retrieval in high dimensional multimedia databases", in *TR-MARS-10-97, 1997*.

86. R. O. Duda and P. E. Hart, "Pattern Classification and Scene Analysis", Chap. 6, Wiley, New York, 1973.

87. H. J. Zhang and D. Zhong, "A scheme for visual feature based image retrieval", in Proc. SPIE Storage and Retrieval for Image and Video Databases, 1995.

88. 3D Medical Informatics, Information Science in Multiple Dimensions chapter12 pp 333-347, Terry S. Yoo

89. Jayant Mishra, Anubhav Sharma and Kapil Chaturvedi, "An Unsupervised Cluster-based Image Retrieval Algorithm using Relevance Feedback", International Journal of Managing Information Technology (IJMIT) Vol.3, No.2, May 2011

90. Agrawal, Rakesh, Gehrke, Johannes, Gunopulos, Dimitrios, Raghavan, Prabhakar, 1998. "Automatic subspace clustering of high dimensional data for data mining applications." In: Proc. ACM SIGMOD, pp. 94–105.

91. Guojun Gan., "Subspace clustering for high dimensional categorical data", May 2003. Talk Given at SOSGSSD.

92. Harsha S. Nagesh., "High performance subspace clustering for massive data sets." Master's thesis, North-western Univeristy, 2145 Sheridan Road, Evanston IL 60208, June 1999.

93. K. G. Woo and J. H. Lee. FINDIT: "a Fast and Intelligent Subspace Clustering Algorithm using Dimension Voting". PhD thesis, Korea Advanced Institute of Science and Technology, Taejon, Korea, 2002.

94. Chun-Hung Cheng, Ada Waichee Fu, and Yi Zhang., "Entropy-based subspace clustering for mining numerical data", In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 84{93. ACM Press, 1999.