INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume3, Issue1)
Available online at: www.ijariit.com

A Survey on Clustering Algorithms used to Perform Image Segmentation

Sandhya Prabhakar H
M.Tech Student, Dept. of CSE
Cambridge Institute of Technology
Bangalore, Karnataka
sandhyahp.86@gmail.com

Mr. Sandeep Kumar
Associate Professor, Dept. of CSE
Cambridge Institute of Technology
Bangalore, Karnataka
sandeep.cse@citech.edu.in

Abstract: The goal of this survey is to use different Clustering techniques to perform image segmentation. Clustering means grouping of images which share some common attributes. The purpose of clustering is to get meaningful result, effective storage and fast retrieval in various areas. The clustering methods are mainly divided into: hierarchical, partitioning, density-based, model-based, grid-based, and soft-computing methods. The goal of this survey is to provide a comprehensive review of different clustering techniques. There are number of clustering algorithms proposed to perform image segmentation. One needs to choose the best algorithm among them by analyzing the nature of the input image in order to get optimal results.

Keywords: Agglomerative, Clustering, Hierarchical, Fuzzy-C, Image Segmentation, k-means.

1. INTRODUCTION

Image segmentation means partitioning the image into different regions which contain similar attributes which may be color, intensity or texture [1-2]. Segmentation is a pre-processing step. It helps in object recognition, tracking and image analysis [15]. Image segmentation results in the collection of segments which combine to form the entire image. There are many issues faced during image segmentation and the main objectives are to minimize overall deviation, maximize connectivity, minimize the error rate, etc.

Clustering means grouping of images which share some common attributes. Clustering can be achieved based on some criterion, so that all the objects that belong to a cluster have some kind of belongingness among them. The specific criterion to be used depends on the application.

The main objectives of clustering algorithms are: scalability, dealing with different types of attributes, discovering clusters with arbitrary shape, minimal requirements of domain knowledge to determine input parameters, ability to deal with noise and outliers, insensitivity to order of input records, high dimensionality and usability. Also there are wide ranges of issues that are faced during clustering like: dealing with large number of dimensions and large number of data items which can become an overhead because of time complexity, defining distance measure should me made appropriately which is a tedious task when it comes to multidimensional spaces, etc. Hence care should be taken while applying the clustering algorithm to images.

2. Classification of Clustering Techniques:

Clustering algorithms may be classified as listed below:

- Exclusive Clustering
- Overlapping Clustering
- Hierarchical Clustering
- Probabilistic Clustering

In the first case data are grouped in an exclusive way, so that if a certain datum belongs to a definite cluster then it could not be included in another cluster. On the contrary the second type, the overlapping clustering, uses fuzzy sets to cluster data, so that each point may belong to two or more clusters with different degrees of membership.

2.1 Hierarchical Clustering Algorithms

Hierarchical clustering algorithms are also known as clustering algorithms through connectivity and have a basic idea that objects are similar to adjacent objects than other more distant objects. These algorithms connect objects to create clusters based on the distance between them. A cluster can be described by the maximum distance needed to connect its parts [26]. The clusters structure can be represented by a dendogram, hence the name - hierarchical clustering [30].

The clustering algorithms family differ by approach and by using different metrics for measuring the distance between objects (Euclidean distance, Manhattan distance, etc.). The user has to choose the criteria to join the two clusters. Some common choices are "single link"(minimum object distance), "complete link" (maximum distances of objects) or "average link" (average distances of objects) [27].

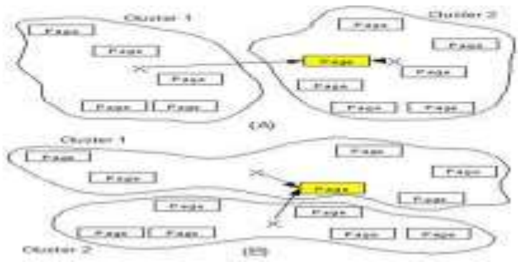

Figure 1: Hierarchical Clustering [48]

Clustering algorithms are classified into two categories: agglomeration and division [25].
Agglomeration algorithms are based on simple elements that create individually one cluster and continue by joining them. On the other hand, division algorithms use, as starting point, a cluster consisting of all elements of the dataset and continue by dividing it into smaller clusters [28].

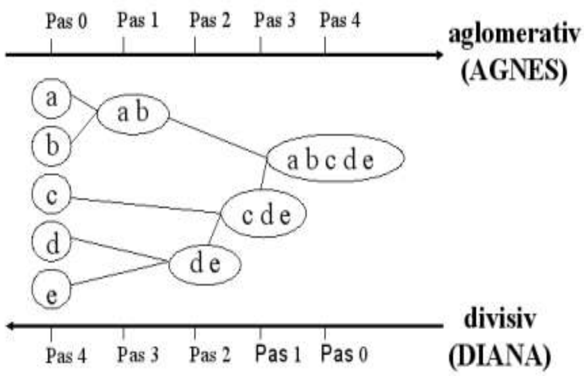

Figure 2: Splitting and clustering hierarchical Agglomerative schema [25]

2.1.1 Agglomerative Nesting

AGNES algorithms have the following steps:
- initialize n clusters, where each object, in the initial set, defines a cluster;
- calculate a triangular matrix of dissimilarity made of distances between all pairs of objects, $d_{i,j} = dist(x_i,x_j)$;
- repeat it by n times (n - the number of dataset elements);
- identify in the dissimilarity matrix the smallest distance between the elements i and j;
- connects i and j clusters to a new one (i,j);
- replace objects i and with (i,j);
- replace the dissimilarity matrix according to the new created way. This steps means to erase the i and j lines and columns and add a new line (i,j) for the distance which will be calculated conforming to the desired link(simple, complex or medium) [26].

2.2.2 Divisive Analysis

Divisive Analysis algorithms have the following steps:
- initialize one single cluster which contains all the dataset objects;
- repeat this for n times;
- find the cluster with the largest diameter (the largest distance between two contained objects);
- find in this cluster the object with the highest dissimilarity reported

for every new object outside the cluster the difference is computed:
Di= [media (d (i, j)), j ∈ new cluster] – [media (d (i,j)),j ∈new cluster]
- find the i object with the largest Di difference; if this is positive, the object will be added to the newest cluster;
- repeat the last two steps until all the Di differences are negative. Now, the cluster was been divided into two new clusters [26].

The algorithm will end when all clusters will contain only one object. In the image classification context of the datasets, when we have a clusters structure generated previously, the problem is no longer when comparing each image from the database, but when comparison with each cluster of images. Thus for each cluster a representative object can be chosen as its center image or can be calculated the centroid's cluster. This is an average of elements characteristics from the cluster. Clusters classification is performed by comparing the queried image with the representative centroid element from the cluster. Further classification can be done by comparing the cluster level with each component image.

Generally, one chooses a distance which seems appropriate for the particular data set. For example, one may choose distance between the closest elements as the inter-cluster distance - which tends to yield extended clusters (statisticians call this method single-link clustering).

Another natural choice is the maximum distance between an element of the first cluster and one of the second - this tends to yield "rounded" clusters (statisticians call this method complete-link clustering).Finally, one could use an average of distances between elements in the clusters - this will also tend to yield "rounded" clusters (statisticians call this method group average clustering). This is an intrinsically difficult task if there is no model for the process that generated the clusters.

2.2 Log Based Relevance Feedback

In a CBIR system, after a user submits a query by a given sample, the system will return a set of similar images to the user. The returned images may not be fully relevant to the user's targets. In order to learn the query concept of the user, relevance feedback is engaged as a query refinement technique for helping the retrieval task. The relevance feed-back mechanism solicits the user to mark the relevance on the retrieved images and then refines the results by learning the feedbacks from the user. The relevance feedback

Procedure is repeated again and again until the targets are found. As stated previously, the semantic gap problem in CBIR is very challenging and regular learning techniques normally need a lot of rounds of feedback for finding satisfactory results. In order to reduce the learning difficulty, one can seek the help from the user's feedback logs which can be available in CBIR systems from a long-term learning perspective.

To engage the users' feedback logs in CBIR, one may refer to the similar techniques studied in traditional information retrieval in which a lot of studies have described how to employ the users' logs to improve the retrieval performance [32, 31]. Some typical approaches are based on the query expansion (QEX) techniques [32, 43, 39, 42, and 37]. Query expansion is viewed as a multiple-instance sampling technique [36] in which the returned samples of the next round are selected from the neighbourhood from the positive samples of the previous rounds. Although query expansion showed successes in document retrieval, it may not be very effective to solve the problem in multimedia retrieval [38, 41].In order to learn the users' logs effectively; we employ a popular yet powerful machine learning technique –support vector machines (SVM)-to attack the problem. SVM is a state-of-the-art classification technique with very good generalization performance applications in relevance feedback [33, 34, 35, 40]. In each relevance feedback session, a user will specify N+ samples are relevant (positive) and N- samples are irrelevant (negative). The information will be logged in the database after the end of the session. When retrieving the relevance feedback information from the logs, a relevance matrix (RM) is constructed to describe the relevance relationship between the images in the database. The column of the relevance matrix represents the image samples in the image database and the row represents the session number in the log database. For each given session, the relevance information of positive samples and negative samples are recorded as relevant (+1), irrelevant (-1) or unknown (0). For example, suppose image i is marked as relevant and j is marked as irrelevant in a given session k, then the corresponding value in the matrix is RM (k; i) = 1 and RM (k; j) = -1. Therefore, relationship of two images i and j can be computed by the following modified correlation formula:

$$R_{ij} = \sum_k \delta_k \cdot RM(k,i) \cdot RM(k,j)$$

$$\delta_k = \begin{cases} 1 & \text{if} \quad RM(k,i) + RM(k,j) \geq 0 \,, \\ 0 & \text{if} \quad RM(k,i) + RM(k,j) < 0 \,. \end{cases}$$

Term is engaged to remove the element pair (-1;-1) for the δ_k correlation formula. If R_{ij} is positive, it indicates that image i and image j are relevant otherwise they are irrelevant. Then for each given image sample, we can find a set of relevant samples and a set of irrelevant samples ranking by their correlation formula. If R_{ij} is positive, it indicates that image i and image j are relevant otherwise they are irrelevant. Then for each given image sample, we can find a set of relevant samples and a set of irrelevant samples ranking by their relationship. From the above descriptions, we can find that the relevant samples for each given sample may be with different relationship values, meaning that their relationship is with different confidence degrees. Typical SVM cannot well explore the relevance information with different confidence degrees. In order to utilize the advantages of SVM for solving the relevance feedback problem, a modified support vector machine called soft label support vector machine is used.

2.4 Fuzzy Clustering

Fuzzy clustering is another technique which found its way as one of the soft means of clustering technique. It's a very recent approach used in many applications to solve diverse problems. They have dynamic capability to improve the segmentation accuracy of different images especially for the real world applications. Motivation of using fuzzy technology is the high degree of adaptability to complex and dynamic nature of segmentation problems. It is mainly used to process the color images [8].It is considered to be soft way of clustering as compared to the K-means clustering technique.

In fuzzy clustering technique, each point has a degree of belongingness to the clusters, rather than belonging completely to just a single cluster. Hence we see that the points on the edges of the cluster may lie in the cluster to a lower degree as compared to the points in the center of cluster. Any point x having a set of coefficients gives larger degree of being in the means; the centroid of the cluster is considered the mean of all the points, weighted by their own degree of belonging to the cluster [16-20].

Fuzzy C-means partitions a set of *n* objects *x* {*x* , *x* ,..., *xn*} in R^d dimensional space into $c(1 < c < n)$ fuzzy clusters with y = { *y1*, *y2* , *y3* ,..., *yc*} cluster centers or centroid [6]. The fuzzy clustering technique of objects is described by a matrix called fuzzy matrix μ with *n* number of rows and *c* number of columns in which *n* is the number of data objects and *c* in the number of clusters. μ_{ij}, the element present in the *i*th row and *j*th column in μ verifies the degree of association or belonging function of the *i*th object with the *j*th cluster. The objective function of Fuzzy C-means algorithm is to minimize the following equation:

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} u_{ij}^{m} d_{ij}$$

Where

$$d_{ij} = \| x_i - y_j \|$$

The Fuzzy C-means algorithm is iterative and can be explained as follows:

1. Select m ($m > 1$); initialize the belonging function values μ_{ij} , $i = 1, 2... n; j = 1,2,...,c$.
2. Calculate the cluster centers yj , $j = 1,2,...,$,c According to above given equation.
3. Compute Euclidian distance d_{ij}, i=1,2…,n

2.5 k-means Clustering

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to recalculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function is a chosen distance measure between a data point and the cluster is an indicator of the distance of the n data points from their respective cluster centers.

The algorithm is composed of the following steps:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Although it can be proved that the procedure will always terminate, the k-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centres. The k-means algorithm can be run multiple times to reduce this effect.

K-means is a simple algorithm that has been adapted to many problem domains. As we are going to see, it is a good candidate for extension to work with fuzzy feature vectors.

2.6 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 [44]. This method can be recursively applied to get more than two clusters. In this method each time the sub graph 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.

2.7 Model-Based Clustering

There's another way to deal with clustering problems: a model based approach, which consists in using certain models for clusters and attempting to optimize the fit between the data and the model. In practice, each cluster can be mathematically represented by a parametric distribution, like a Gaussian (continuous) or a Poisson (discrete). The entire data set is therefore modeled by a mixture of these distributions. An individual distribution used to model a specific cluster is often referred to as a component distribution.

A mixture model with high likelihood tends to have the following traits:

- Component distributions have high "peaks" (data in one cluster are tight).
- The mixture model "covers" the data well (dominant patterns in the data are captured by component distributions).

Advantages of model-based clustering:

- Well-studied statistical inference techniques available.
- Flexibility in choosing the component distribution.
- Obtain density estimation for each cluster.
- A "soft" classification is available.

3. Distance Measures

Since clustering is the grouping of similar instances/objects, some of measure that can determine whether two objects are similar or dissimilar is required. There is main type of measures used to estimate this relation: distance measures and similarity measures. Many clustering methods use distance measures to determine the similarity or dissimilarity between any pair of objects. It is calculated for Numeric Attributes, Binary Attributes, Nominal Attributes, ordinal Attributes, and Mixed-Type Attributes.

4. Applications of Clustering Algorithms

Few areas where Clustering algorithms can be applied:

- Marketing: finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records;
- Biology: classification of plants and animals given their features;
- Libraries: book ordering;
- Insurance: identifying groups of motor insurance policy holders with a high average claim cost; identifying frauds;
- City-planning: identifying groups of houses according to their house type, value and geographical location;
- Earthquake studies: clustering observed earthquake epicenters to identify dangerous zones;
- WWW: document classification; clustering weblog data to discover groups of similar access patterns.

Conclusion

The goal of image segmentation is a domain-independent decomposition of an image into distinct regions. Clustering concepts and image segmentation concepts have been analyzed. Image segmentation has become a very important task in today's scenario. In the present day world computer vision has become an interdisciplinary field and its applications can be found in any area. Thus, to find an appropriate segmentation algorithm and clustering algorithm based on your application and the type of inputted image is very important.

References

[1] Gonzalez, R.C.; Woods, R.E. Digital Imaging Processing; Prentice Hall: New York, NY, USA, 2002.
[2] S. Raut, M. Raghuvanshi, R. Dharaskar, A.Raut,"Image Segmentation-A State-of-Art Survey for Prediction," ICACC, pp.420-424,2009.

[3] S. Naz, H. Majeed, H. Irshad,"Image segmentation using fuzzy clustering: A survey," International conference on emerging technologies, pp.181- 186,2010.

[4] L. Wang,"Comparison for Edge Detection of Colony
Image," IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.9A,2006.
[5] Z. Musoromy, R. Soodamani, B. Nico,"Edge Detection Comparison for License Plate Detection,"Int.Control, Automation, Robotics and Vision,2010.
[6] I. Karoui, J. Boucher, J. Augustin,"Variationalbegion-Based Segmentation Using Multiple Texture Statistics," IEEE trans on Image processing, vol.19,no.12,2010..
[7] C. Rambabu, I. Chakrabarti, A. Mahanta,"Floodingbased watershed algorithm and its prototype hardware architecture," IEEE Proceedings vision image and signal processing, vol.151,no.3,pp.224-234,2004.
[8] Ahmed A.Othman,Hamid R. Tiszhoosh,"Evolving Fuzzy Image Segmentation", IEEE Proceedings ICFS, pp1603-1608,2011.
[9] N. A. Mat-Isa, "Automated edge detection technique for Pap Smear images using moving k-means clustering and modified seed based region growing algorithm," International Journal of the Computer, the Internet and Management, vol. 13, pp. 45-59, 2005.
[10] N. A. Mat-Isa, M. Y. Mashor, and N. H. Othman,"Comparison of segmentation performance of clustering algorithms for Pap smear images". Proceedings of International Conference on Robotics,Vision, Information and Signal processing (ROVISP2003).pp. 118-125, 2003.
[11] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A.Y. Wu "An efficient k-means clustering algorithm: analysis and implementation," IEEE Trans.on Pattern Analysis and machine Intelligence, vol. 24, No. 7, 2002.
[12] L. Cinque, G. Foresti, and L. Lombardi, "A clustering fuzzy approach for image segmentation," Pattern Recognition, vol. 37, no. 9, pp. 1797–1807, 2004.
[13] H. Zhang, J.E. Frittsm, S.A. Goldman, "Image segmentation evaluation: a survey of unsupervised methods," Computer Vision and Image Understanding, Vol. 110 (2), pp. 260-280, 2008.
[14] K. Mc Guinness, and N. O'Connor,"A comparative evaluation of interactive segmentation algorithms," Pattern Recognition, Vol. 43, pp.434-444, 2010.
[15] N. R. Pal, and S. K. Pal, "A review on image segmentation techniques, "Pattern Recognition, vol. 26, pp. 1277–1294, September 1993.
[16] M. C. Hung and D. L. Yang, "An efficient fuzzy cmeans clustering algorithm," in Proc. IEEE Int. Conf. Data Mining, 2001, pp. 225–232.
[17] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms. New York, NY, USA:Plenum, 1981.
[18] P. S. Bradley and U. M. Fayyad, "Refining initial points for k-means clustering," Microsoft Res., Redmond, WA, USA, Tech. Rep. MSR-TR- 98-36, May 1998.
[19] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," IEEE Trans. Fuzzy Syst., vol. 3, no. 3, pp. 370–379, Aug. 1995.
[20] C. Hwang, F. C. H. Rhee, "Uncertain Fuzzy clustering: Interval Type-2 Fuzzy Approach to CMeans", IEEE Tran. on Fuzzy Systems, Vol. 15, 107- 120, 2007.
[21] Yixin Chen, James Z. Wang, Robert Krovetz. "Content-Based Image Retrieval By Clustering", MIR '03 "Proceedings of the 5th ACM SIGMM international workshop on Multimedia information retrieva", 2003.
[22] John Eakins, Margaret Graham "Content-based Image Retrieval", The JISC Technology Applications Programme", 1999.
[23] M. Pietikainen, G. Zhao, A. Hadid, T. Ahonen, "Computer Vision Using Local Binary Patterns", Springer-Verlag London Limited,
2011.
[24] Xiang-Yang Wang, Jun-Feng Wu, Hong-Ying Yang "Robust image retrieval based on color histogram of local feature regions", Springer Netherlands, 2009.
[25] Jiawei Han, MichelineKamber, "Data Mining - Concepts and Techniques", 2nd edition, Morgan-Kaufmann, 2006.
[26] Kaufman, L.; Rousseeuw, P.J. (1990)." Finding Groups in Data: An Introduction to Cluster Analysis" (1 ed.). New York: John Wiley.
[27] P.-N. Tan, M. Steinbach & V. Kumar, "Introduction to Data Mining", , Addison-Wesley (2005).
[28] Bailey, Ken, "Numerical Taxonomy and Cluster Analysis". Typologies and Taxonomies, 1994.
[29] James Z. Wang, Jia Li, GioWiederhold, ``SIMPLIcity: Semantics-sensitive Integrated Matching for Picture
LIbraries," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 23, no.9, pp. 947-963, 2001.
[30] Xindong Wu, Vipin Kumar, J. Ross Quinlan, JoydeepGhosh, Qiang Yang, Hiroshi Motoda, Geoffrey J. McLachlan, Angus Ng, Bing Liu, Philip S. Yu, Zhi-Hua Zhou, Michael Steinbach, David J. Hand, and Dan Steinberg. 2007. Top 10 algorithms in data mining. Knowl. Inf. Syst.14, 1 (December 2007), 1-37
[31] P. Anick. Using terminological feedback for web search refinement: a log-based study.In Proceedings of the 26th Intl. ACM SIGIR Conf. (SIGIR'03), pages 88-95, 2003.
[32] H. Cui, J.-R. Wen, J.-Y.Nie, and W.-Y. Ma. Probabilistic query expansion using query logs. In Proceedings of the 11th Intl. World Wide Web Conf. (WWW'2002), 2002.
[33] C.-H. Hoi and M. R. Lyu. Biased support vector machine for relevance feedback in image retrieval. In Proceedings of Intl. Joint Conf. on Neural Networks (IJCNN'04), Budapest, Hungary, 2004.
[34] C.-H. Hoi and M. R. Lyu.Group-based relevance feeedback with support vector machine ensembles.In Proceedings of the 17th Intl. Conf. on Pattern Recognition (ICPR'04), Cambridge, UK, 2004.
[35] P. Hong, Q. Tian, and T. Huang. Incorporate support vector machines to content-based image retrieval with relevant feedback. In Proceedings of IEEE Intl. Conf. on Image Processing (ICIP'00), volume 3, pages 750 -753, Vancouver, BC, Canada, 2000.
[36] Y. Ishikawa, R. Subramanya, and C. Faloutsos.MindReader: Querying databases through multiple examples. In Proceedings of the 24th Int. Conf. Very Large Data Bases (VLDB'98), pages 218-227, 1998.

[37] K. Porkaew, K. Chakrabarti, and S. Mehrotra. Query refinement for multimedia retrieval and its evaluation techniques in mars. In Proceedings of ACM Intl. Conf. on Multimedia (MM'99), Orlando, Florida, USA, 1999.
[38] K. Porkaew, M. Ortega, and S. Mehrotra. Query reformulation for content based multimedia retrieval in MARS. In Proceedings of ICMCS, volume 2, pages 747-751, 1999.
[39] G. Salton and C. Buckley.Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science, 44(4):288-287, 1990.
[40] S. Tong and E. Chang. Support vector machine active
learning for image retrieval. In Proceedings of the 9th ACM Intl.Conf.on Multimedia, pages 107-118. ACMPress, 2001.
[41] V. N. Vapnik. Statistical Learning Theory. Wiley,1998.
[42] L. Wu, C. Faloutsos, K. P. Sycara, and T. R. Payne.FALCON: Feedback adaptive loop for content-based retrieval. The VLDB Journal, pages 297{306, 2000.
[43] J. Xu and W. B. Croft. Query expansion using local and global document analysis. In Proceedings of the 19th Intl. ACM SIGIR Conf. (SIGIR'96), pages 4{11, 1996.
[44] R. Kannan, S. Vempala, A. Vetta, "On Clustering -good, bad and spectral", FOCS, 2000, pp. 367-377.
[45]Neha Mathur; Pankaj Dadheech; Mukesh Kumar Gupta," The K
means Clustering Based Fuzzy Edge Detection Technique on MRI Images", 2015
[46]Radu Andrei Stefan; Ildikó-Angelica Szöke; Stefan Holban,
"Hierarchical clustering techniques and classification applied in ContentBased Image Retrieval (CBIR) ",(2015).
[47] Shraddha Gangwar; Rishi Pal Chauhan, "Survey of Clustering Techniques Enhancing Image Segmentation Process", (2015).
[48] M. Lalitha1, M. Kiruthiga2, C. Loganathan3, "A Survey on Image Segmentation through Clustering Algorithm", (2013).
[49] Chu-Hong Hoi and Michael R. Lyu Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong S.A.R, "A Novel Log-based Relevance Feedback Technique in Content-based Image Retrieval".
[50]750-753, Vancouver, BC, Canada, 2000.
http://www.cs.bris.ac.uk/home/tr1690/documentation/fuzzy_clustering_initial_report/node11.html
[51] Osmar R. Zaïane: "Principles of Knowledge Discovery in Databases – Chapter8: Data Clustering"
http://www.cs.ualberta.ca/~zaiane/courses/cmput690/slides/Chapter8/index.html
[53] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability"*, Berkeley, University of California Press, 1:281-297
[54] Andrew Moore: "K-means and Hierarchical Clustering - Tutorial Slides"
http://www2.cs.cmu.edu/~awm/tutorials/kmeans.html
[55] Brian T. Luke: "K-Means Clustering"
http://fconyx.ncifcrf.gov/~lukeb/kmeans.html
[56] Tariq Rashid: "Clustering"
http://www.cs.bris.ac.uk/home/tr1690/documentation/fuzzy_clustering_initial_report/node11.html
[57] Hans-Joachim Mucha and HizirSofyan: "Nonhierarchical Clustering"
http://www.quantlet.com/mdstat/scripts/xag/html/xaghtmlframe149.ht
[58] Osmar R. Zaïane: "Principles of Knowledge Discovery in Databases - Chapter 8: Data Clustering"
http://www.cs.ualberta.ca/~zaiane/courses/cmput690/slides/Chapter8/index.html
[59] Jia Li: "Data Mining - Clustering by Mixture Models"
http://www.stat.psu.edu/~jiali/course/stat597e/notes/mix.pd

