Multi-Dimensional Features Reduction of Consistency Subset Evaluator on Unsupervised Expectation Maximization Classifier for Imaging Surveillance Application

Chue-Poh TAN chue.poh@mimos.my
Centre for Advanced Informatics
MIMOS Berhad
Technology Park Malaysia
57000 Kuala Lumpur, Malaysia.

Ka-Sing LIM kslim@mmu.edu.my
Faculty of Engineering,
Multimedia University,
Persiaran Multimedia,
63100 Cyberjaya, Selangor, Malaysia.

Weng-Kin LAI (Dr.) weng.kin@mimos.my
Centre for Advanced Informatics
MIMOS Berhad
Technology Park Malaysia
57000 Kuala Lumpur, Malaysia.

Abstract

This paper presents the application of multi dimensional feature reduction of Consistency Subset Evaluator (CSE) and Principal Component Analysis (PCA) and Unsupervised Expectation Maximization (UEM) classifier for imaging surveillance system. Recently, research in image processing has raised much interest in the security surveillance systems community. Weapon detection is one of the greatest challenges facing by the community recently. In order to overcome this issue, application of the UEM classifier is performed to focus on the need of detecting dangerous weapons. However, CSE and PCA are used to explore the usefulness of each feature and reduce the multi dimensional features to simplified features with no underlying hidden structure. In this paper, we take advantage of the simplified features and classifier to categorize images object with the hope to detect dangerous weapons effectively. In order to validate the effectiveness of the UEM classifier, several classifiers are used to compare the overall accuracy of the system with the compliment from the features reduction of CSE and PCA. These unsupervised classifiers include Farthest First, Density-based Clustering and k-Means methods. The final outcome of this research clearly indicates that UEM has the ability in improving the classification accuracy using the extracted features from the multi-dimensional feature reduction of CSE. Besides, it is also shown that PCA is able to speed-up the computational time with the reduced dimensionality of the features compromising the slight decrease of accuracy.

Keywords: Consistency Subset Evaluator, Principal Component Analysis, Unsupervised Expectation Maximization, Classification, Imaging surveillance

1. INTRODUCTION

Security surveillance systems are becoming indispensable in scenarios where personal safety could be jeopardized due to criminal activities [1]. Conventional security surveillance systems require the constant attention of security personnel, who monitor several locations concurrently [2,3]. Hence, the advancement in image processing techniques has become an advantage to the security surveillance systems to improve on the operational activity for monitoring purpose.

Image classification is an essential process in image processing and its major issue lies in categorizing images with huge number of input features using traditional classification algorithm. These algorithms tend to produce unstable prediction models with low generalization performance [4]. To overcome high dimensionality, image classification usually relies on a pre-processing step, specifically to extract a reduced set of meaningful features from the initial set of huge number of input features. Recent advances in classification algorithm have produced new methods that are able to handle more complex problems.

In this paper, we emphasize on the analysis and usage of the multi-dimensional features reduction on advanced classification method of Unsupervised Expectation Maximization (UEM) to classify dangerous weapons within an image. In order to validate the effectiveness of the feature reduction method and classifier, several classifiers such as Farthest First, Density-based Clustering and k-Means methods are utilized to compare the overall accuracy of the classifiers. Finally, the study depicts the comparative analysis of different classification techniques with respect to the robustness of the meaningful extracted features. The classification process comprised of four steps, which are feature extraction, training, prediction and assessing the accuracy of the classification. Analysis on the features is done to ensure the robustness and usefulness of each feature to differentiate classes effectively. The details of the classification will be discussed in this paper.

This paper is divided into four sections. Section II presents the methodology and the dataset used in this paper. In this section, the basic concept of Consistency Subset Evaluator (CSE), Principal Component Analysis (PCA), Expectation Maximization (UEM), Farthest First, Density-based Clustering and k-Means methods are discussed. Section III describes the results and discussion for the findings of the classification process using the aforementioned classifiers. The accuracy assessment with the comparisons between the classifiers is discussed in this section. In Section IV, we conclude this paper with the suggestion on future work.

2. METHODOLOGY

2.1 Data Description

In this paper, we utilized on a set of data which was available freely in the internet [5] to carry out some experimental research on the classification. We evaluated the selected algorithms using the training dataset which contains 13 features (attributes value of the image objects) with their associate class labels (Human, Bull, Child, Dog, Duck, Knife classes). Besides, 6 test dataset that contain the same features value of the image objects for each class have been identified. Feature extraction process was carried out to extract all useful features from 128 binary images (black and white images) to represent the characteristics of the image object. From the image analysis and feature extraction, 13 important and useful features of the image object as the attributes of the dataset were obtained. In this case, the extracted features must be robust enough and RST (rotation, scale and transition) invariant. A very adaptive feature would be RST-invariant, meaning that if the image object is rotated, shrunk or enlarge or translated, the value of the feature will not change. We took the invariance of each feature into consideration and the features comprised of compactness, elongation, ratio of major axis length and minor axis length, hull ratio, moment, area ellipse ratio, axis ratio, ratio between area of the bounding box minus area of the blob and area of the bounding box, ratio between the height and the width of the bounding box, ratio between the squared perimeter and the area of the blob, roughness, ratio of the area of the blob and the area of the bounding box and compactness circularity of the blob.

2.2 Multi-dimensional Feature Reduction Methods

Feature reduction process can be viewed as a preprocessing step which removes distracting variance from a dataset, so that classifiers can perform better. In this paper, we present two multi-dimensional feature reduction methods, namely Consistency Subset Evaluator (CSE) and Principal Component Analysis (PCA).

2.2.1 Consistency Subset Evaluator (CSE)

Class consistency has been used as an evaluation metric by several approaches to attribute subset evaluation [6-8]. Attribute subset evaluation is done to look for combinations of attributes whose values divide the data into subsets containing a strong single class majority [9]. The search is in favor of small feature subsets with high class consistency. This consistency subset evaluator uses the consistency metric presented by H. Liu et al. as shown in Equation (1)

$$Consistency_s = 1 - \frac{\sum_{i=0}^{J} |D_i| - |M_i|}{N} \tag{1}$$

where s is an attribute subset, J is the number is the number of distinct combinations of attribute values for s, $|D_i|$ is the number of occurrences of the ith attribute value combination, $|M_i|$ is the cardinality of the majority class for the ith attribute value combination and N is the total number of instances in the data set [9].

To use the Consistency Subset Evaluator, the dataset needs to be discretized with numeric attributes using any suitable method such as the method of U. M. Fayyad et al. [10]. The search method that can be used is the forward selection search which is to produce a list of attributes [11]. The attributes are then ranked according to their overall contribution to the consistency of the attribute set.

2.2.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) is one of the most popular multi dimensional features reduction products derived from the applied linear algebra. PCA is used abundantly because it is a simple and non-parametric technique of extracting relevant information from complex data sets. The goal of PCA is to reduce the dimensionality of the data while retaining as much as possible of the variation in the original dataset.

Suppose x_1, x_2, … x_N are *N* x 1 vectors.

Step 1: Mean value is calculated with Equation (2).

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2}$$

Step 2: Each features is used to subtract the mean value, shown in Equation (3).

$$\Phi_i = x_i - \bar{x} \tag{3}$$

Step 3: Matrix $A = [\Phi_1 \quad \Phi_2 \quad \mathrm{K} \quad \Phi_N]$ is generated with *N* x *N* matrix and covariance matrix with the same dimension size is computed as Equation (4) [12].

$$C = \frac{1}{M}\sum_{i=1}^{N} \Phi_i \Phi_i^T = AA^T \tag{4}$$

The covariance matrix characterizes the distribution of the data.
Step 4: Eigenvalues is computed:

$$C = \lambda_1 > \lambda_2 > \ldots > \lambda_N \tag{5}$$

Step 5: Eigenvector is computed:

$$C = u_1, u_2, \ldots, u_N \tag{6}$$

Since C is symmetric, $u_1, u_2, \ldots, u_N$ form a basis, ($(x - \bar{x})$, can be written as a linear combination of the eigenvectors):

$$x - \bar{x} = b_i u_1 + b_2 u_2 + \ldots + b_N u_N = \sum_{i=1}^{N} b_i u_i \tag{7}$$

Step 6: For dimensionality reduction, it keeps only the terms corresponding to the K largest eigenvalues [13]

$$x - \bar{x} = \sum_{i=1}^{K} b_i u_i \quad \text{where } K << N \tag{8}$$

The representation of x into the basis $u_1, u_2, \ldots, u_K$ is thus

$$\begin{bmatrix} b_i \\ b_2 \\ \ldots \\ b_K \end{bmatrix}. \tag{9}$$

2.3 Classification Methods

The aim is to do comparison of supervised classification methods for classification of the image object to their known class from the reduced multi-dimensional features dataset. The issue in identifying the most promising classification method to do pattern classification is still in research. Therefore, we are interested in predicting the most promising classification method for pattern classification in terms of the classification accuracy achieved in detecting dangerous weapons. The algorithms considered in this study are UEM, Farthest First, Density-based Clustering and k-Means. The methodology for each classifier is presented with basic concept and background.

2.3.1 Unsupervised Expectation Maximization (UEM)

The algorithm is in a model-based methods group which hypothesizes a model for each of the clusters and finds the best fit of the data to the given model [14]. Expectation Maximization performs the unsupervised classification or learning based on statistical modeling [15].
A cluster can be represented mathematically by a parametric probability distribution

$$\mathrm{P}(x_i \in C_k) = p(C_k \mid x_i) = \frac{p(C_k)p(x_i \mid C_k)}{p(x_i)} \tag{10}$$

where each object x_i is assigned to cluster C_k and $p(x_i \mid C_k) = N(m_k, E_k(x_i))$ follows the normal distribution around mean, m_k , with expectation, E_k [16]. The entire data is a mixture of these distributions where each individual distribution is typically referred to as a component distribution which makes use of the finite Gaussian mixture models. So, clustering the data can be done by using a finite mixture density model of k probability distribution [17].

This algorithm can be used to find the parameter estimates for the probability distribution. It assigns each object to a cluster according to a weight representing the probability of membership [16]. Basically the algorithm consists of two main steps which are the Expectation step and the Maximization step. The Expectation step calculates the probability of cluster membership of object x_i, for each cluster and these probabilities are the expected cluster membership for object x_i. On the other hand, the Maximization step uses the probability estimates to re-estimate the model parameters. The Expectation step can be interpreted as constructing a local lower-bound to the posterior distribution, whereas the Maximization step optimizes the bound, thereby improving the estimate for the unknowns [18]. The parameters found on the Maximization step are then used to begin another Expectation step, and the process is repeated [19].

2.3.2 Farthest First Classifier

Farthest First is a unique clustering algorithm that combines hierarchical clustering and distance based clustering. It uses the basic idea of agglomerative hierarchical clustering in combination with a distance measurement criterion that is similar to the one used by K-Means. Farthest-First assigns a center to a random point, and then computes the *k* most distant points [20].

This algorithm works by first select an instance to be a cluster centroid randomly and it will then compute the distance between each remaining instance and its nearest centroid. The algorithm decides that the farthest instance away from its closed centroid as a cluster centroid. The process is repeated until the number of clusters is greater than a predetermined threshold value [21].

2.3.3 Density-based Clustering Classifier

Density based algorithms typically regard clusters as dense regions of objects in the data space that are separated by regions of low density [22]. The main idea of density-based approach is to find regions of low and high density. A common way is to divide the high dimensional feature space into density-based grid units. Units containing relatively high densities are the cluster centers and the boundaries between clusters fall in the regions of low-density units [23].

This method of clustering also known as a set of density-connected objects that is maximal with respect to density-reachability [22]. Regions with a high density of points depict the existence of clusters while regions with a low density of points indicate clusters of noise or clusters of outliers. For each point of a cluster, the neighbourhood of a given radius has to contain at least a minimum number of points, which is, the density in the neighbourhood has to exceed some predefined threshold. This algorithm needs three input parameters, which comprised of the neighbour list size, the radius that delimitate the neighbourhood area of a point, and the minimum number of points that must exist in the radius that deliminate the neighborhood area of a point [24].

2.3.4 K-Means Classifier

K-Means is one of the simplest unsupervised learning algorithms that solve clustering problem. K-Means algorithm takes the input parameter and partitions a set of n objects into k clusters so that the resulting intracluster similarity is high but the intercluster similarity is low [25]. Cluster similarity is measured in regard to the mean value of the object in a cluster which can be viewed as the centroid of the cluster.

The k-Means algorithm randomly selects k of the objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is the most similar based on the distance between the object and cluster mean. Then, it computes the new mean for each cluster and this process iterates until the criterion function converges [26]. The algorithm works well when the clusters are compact clouds that are rather well separate from one another. The method is relatively scalable and efficient in processing large data sets because the computational complexity of the algorithm [27-28].

3. RESULTS AND DISCUSSION

In this study, before any classification is applied on the dataset, CSE and PCA are used to explore the usefulness of each feature and reduce the multi dimensional features to simplified features with no underlying hidden structure. The distributions of each feature are drawn and analyzed statistically. Figure 1 shows the distributions for the features which are discarded after CSE implementation. These features include ratio of major axis length and minor axis length, ratio between the squared perimeter and the area of the blob and ratio of the area of the blob and the area of the bounding box. On the other hand, Figure 2 shows the distributions for the features which are discarded after PCA implementation and these features comprised of hull ratio, axis ratio, ratio between area of the bounding box minus area of the blob and area of the bounding box, ratio of the area of the blob and the area of the bounding box and compactness circularity of the blob

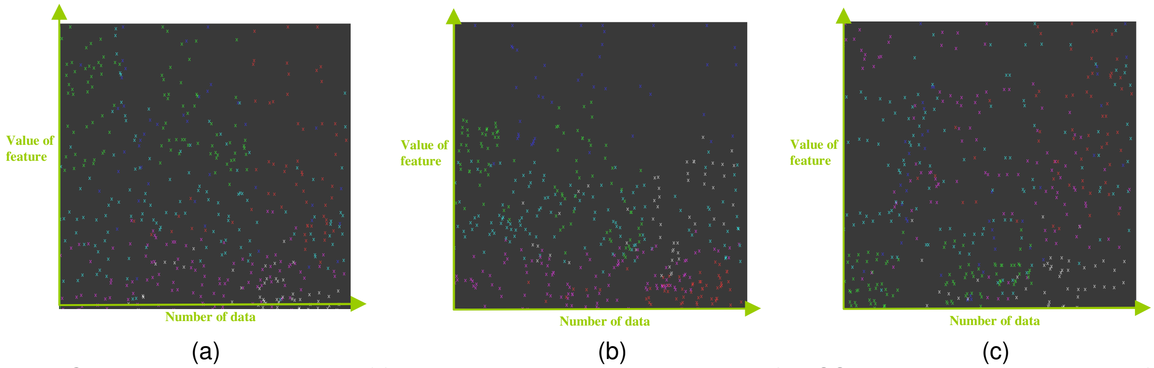

FIGURE 1: The distributions of features which are being discarded after CSE implementation (a) ratio of major axis length and minor axis length, (b) ratio between the squared perimeter and the area of the blob and (c) ratio of the area of the blob and the area of the bounding box

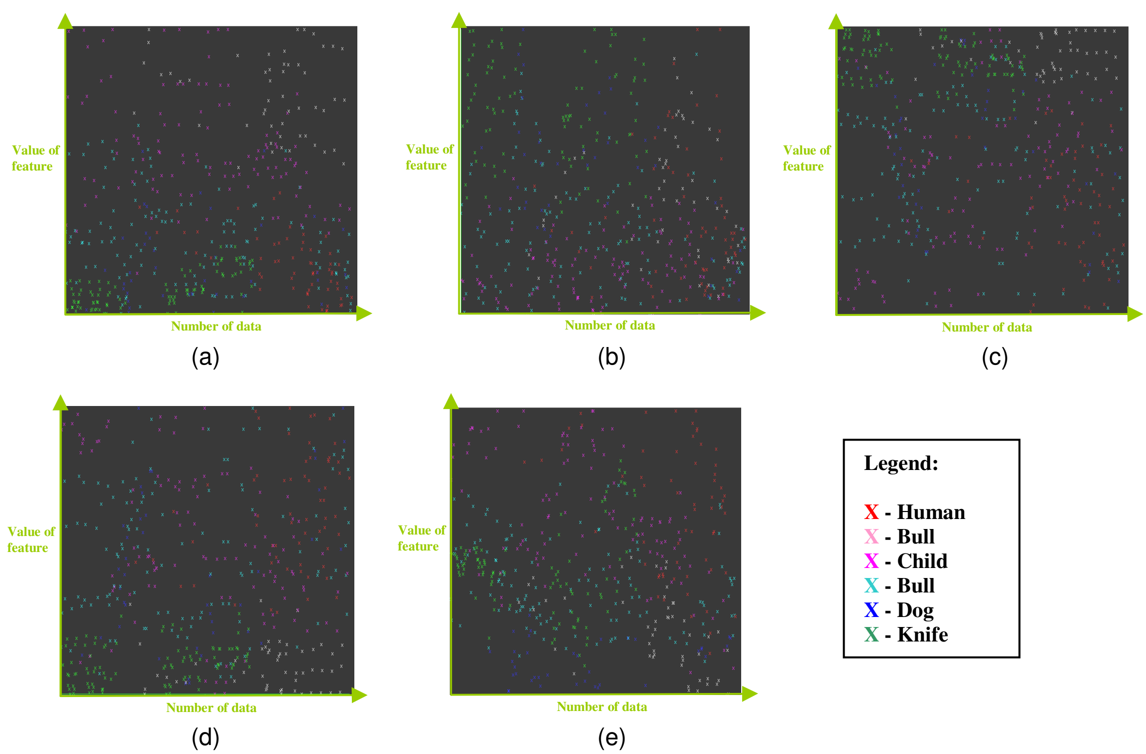

FIGURE 2: The distributions of features which are being discarded after PCA implementation (a) hull ratio, (b) axis ratio, (c) ratio between area of the bounding box minus area of the blob and area of the bounding

box, (d) ratio of the area of the blob and the area of the bounding box and (e) compactness circularity of the blob

The unsupervised classification algorithms, including UEM, Farthest First, Density-based Clustering, and k-Means classifiers are applied to the datasets. In order to validate the impact of multi dimensional feature reduction methods of CSE and PCA, four types of dataset are utilized , namely the original data, data produced after CSE method, data produced after PCA method and data produced after CSE and PCA methods. The classifiers are analyzed and the accuracy assessment is as shown in Table 1 with the computational speed (shown in bracket). In this study, the model with the highest classification accuracy is considered as the best model for pattern classification of this dataset.

	Original data (13 features)	CSE + Classifier	PCA + Classifier	CSE + PCA + Classifier
Expectation Maximization	93.33 % (8.12ms)	95 .83% (7.18ms)	90.12 % (6.21)	92.29 % (4.88ms)
Farthest First	81.88 % (7.33ms)	83.54 % (6.09ms)	82.08 % (5.65ms)	86.25 % (4.26ms)
Density based Clusterer	85.21 % (8.35ms)	88.33% (7.27ms)	87.71 % (6.51ms)	80.21% (4.93ms)
K-Means	86.04 % (7.45ms)	86.88 % (6.15ms)	89.38 % (5.69ms)	81.67% (4.37ms)

Table 1 : Accuracy Assessment and Computational Speed of Experimental Methods on Different Datasets

Based on Table 1, we can see that CSE + UEM classifier achieve the highest overall classification accuracy of all the different datasets. As the dataset we used in this study is quite small and based on our research, UEM classifier is best applied to small dataset. On the other hand, the classifiers with features generated from PCA provide slightly less accuracy and computational speed compared to the classifiers using the predefined number of features. This is due to the reduced dimensional features offered by PCA which allow only the useful key features to participate in the classification process.

4. CONCLUSION

The project is aimed to investigate the performance and impact of CSE and PCA on classification in the aspect of accuracy and computational speed. The potential of each classifier has been demonstrated and the hybrid method of CSE and UEM has shown a desirable result in detecting weapons compared to other classifiers. Our future work shall extend this work to multiple type of images and real-time signal data.

5. REFERENCES

1. A.T. Ali, and E.L. Dagless. *"Computer vision for security surveillance and movement control"*, IEE Colloquium on Electronic Images and Image Processing in Security and Forensic Science, pp. 1-7. 1990.
2. A.C.M. Fong. *"Web-based intelligent surveillance systems for detection of criminal activities"*, Journal of Computing and Control Engineering, 12(6), pp. 263-270, 2001.

3. Y. T. Chien, Y. S. Huang, S. W. Jeng, Y. H. Tasi, and H. X. Zhao. *"A real-time security surveillance system for personal authentication"*. IEEE 37th Annual 2003 International Carnahan Conference on Security Technology 2003, pp. 190-195, 2003.
4. P. Geurts. "*Contribution to decision tree induction: bias/ variance tradeoff and time series classification*". PhD. Thesis, Department of Electrical Engineering and Computer Science, University of Liege, May 2002.
5. http://www.cs.sdce.edu/ShapeMatcher/.
6. H. Almuallim and T. G. Dietterich. *"Learning with many irrelevant features".* Proceedings of the Ninth National Conference on Artificial Intelligence, pp. 547-552, 1991.
7. H. Liu and R. Setiono. *"A probabilistic approach to feature selection".* Proceedings of the 13th International Conference on Machine Learning. pp. 319-327, 1996.
8. M. A. Hall and G. Holmes. *"Benchmarking Attribute Selection Techniques for Discrete Class Data Mining".* IEEE Transactions On Knowledge And Data Engineering, 15(3), 2003.
9. I. Kononenko. *"Estimating attributes: Analysis and extensions of relief".* Proceedings of the Seventh European Conference on Machine Learning, pp. 171-182, 1994.
10. U. M. Fayyad and K. B. Irani. *"Multi-interval discretisation of continuous-valued attributes".* Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, pp. 1022-1027, 1993.
11. M. A. Aizerman, E. M. Braverman, and L.I. Rozoner. *"Theoretical foundations of the potential function method in pattern recognition learning"*. Automation and Remote Control, 25:826-837, 1964.
12. M. E. Tipping and C. M. Bishop. *"Mixtures of probabilistic principal component analyzers".* Neural Computation, 11(2): 443-482, 1999.
13. M. E. Tipping and C. M. Bishop. *"Probabilistic principal component analysis"*. Journal of the Royal Statistical Society, 61(3): 611, 1999.
14. A. Basilevsky. *"Statistical Factor Analysis and Related Methods".* Wiley, New York, 1994.
15. J. Han and M. Kamber. *"Data Mining: Concepts and Techniques".* Morgan Kaufmann, San Francisco, CA (2001).
16. I. Borg and P. Groenen. *"Modern Multidimensional Scaling: Theory and Applications".* Springer (1997).
17. T. F. Cox and M. A. A. Cox. *"Multidimensional Scaling".* Chapman and Hall (2001).
18. S. T. Roweis and L. K. Saul. *"Nonlinear dimensionality reduction by locally linear embedding".* Science, 290(2):2323-2326, 2000.
19. F. Dellaert. *"The Expectation Maximization Algorithm, College of Computing, Georgia Institute of Technology".* Technical Report, 2002.
20. S. D.Hochbaum and B. D. Shmoys. *"A Best Possible Heuristic for the k-Center Problem".* Mathematics of Operational Research, 10(2): pp. 180-184,1985.
21. S. Dasgupta and P. M. Long. *"Performance guarantees for hierarchical clustering".* Journal of Computer and System Sciences, 70(4):555-569, 2005.
22. X. Zheng, Z. Cai and Q. Li. *"An experimental comparison of three kinds of clustering algorithms".* IEEE International Conference on Neural Networks and Brain, pp. 767 -771, 2005.
23. M. Rehman and S. A. Mehdi. *"Comparison of density-based clustering algorithms".* Lahore College for Women University, Lahore, Pakistan, University of Management and Technology, Lahore, Pakistan.
24. M. Ester, H. P. Kriegel, J. Sander and X. Xu, *"A density-based algorithm for discovering clusters in large spatial databases with noise"*. The 2nd International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, USA, 1996.
25. K. Alsabti, S. Ranka and V. Singh. *"An efficient k-Means clustering algorithm".* Available online at http://www.cise.ufl.edu/~ranka/1997.
26. T. H. Cormen, C. E. Leiserson and R. L. Rivest. *"Introduction to algorithms".* McGraw-Hill Book Company, 1990.
27. R. C. Dubes and A. K. Jain. *"Algorithms for clustering data".* Prentice Hall (1998).

28. T. Zhang, R. Ramakrishnan and M. Livny. "An efficient Data Clustering Method for very large databases". Proceedings of the 1996 ACM Sigmod International Conference on Management of Data, Montreal, Canada, pp.103-114, 1996.