Evaluating Clustering Methods for Classification of Marble Slabs in an Automated Industrial Marble Inspection System

M. Alper Selver and Olcay Akay

Dept. of Electrical and Electronics Engineering, Dokuz Eylül University, Izmir, Turkey.
alper.selver@deu.edu.tr, olcay.akay@deu.edu.tr

Abstract

Marble quality classification is a very important issue since export of marble slabs has important economic impact on natural stone industry. Considering the massive transportation costs and requirements of marbles, it is necessary to achieve acceptably high performance in quality classification. However, classification of marble slabs in terms of quality is generally performed manually using human operators. This is rather subjective and prone to errors. Hence, automated and computerized methods are needed in order to obtain reproducible and objective results. In this work, we test the performance of clustering based classification on texture based feature sets. This is done in order to compare performances of different clustering techniques, which might be employed instead of neural networks, especially when there is not enough number of samples to construct a training set. Different clustering methods have been applied to our data set and the results are evaluated by using several cluster validation techniques.

1. Introduction

Export of marble slabs has important economic impact on natural stone industry. Considering the massive transportation costs and requirements of marbles, it is imperative to perform highly accurate classification with respect to quality and appearance [1]. The quality classification process is mostly carried out at the end of the production line, where human experts evaluate and classify the product visually. However, using human experts for classification can be error prone owing to subjective criteria of the operator and the visual fatigue after a period of time, which together degrade the classification performance. Thus, it is necessary to use an automated system capable of performing the same classification tasks that are currently carried out by human experts. For these purposes, a new electro-mechanical system, which automatically classifies marble slabs while they are on a conveyor belt and groups them with the help of a control mechanism, has recently been proposed. The developed system is composed of two parts: The software part acquires marble images, extracts several features, and finally classifies them using Artificial Neural Networks (ANN). Hardware part is composed of a conveyor belt, a serial port communication system, pneumatic pistons, programmable logic controller and their control circuits for grouping the marbles mechanically. Prior to this design, a very large and diverse data set containing 1158 marble surface images (193 marble cube specimens times 6 surfaces), which are classified into four quality groups (shown in Figure 1) by human experts, had been constructed. Then, the above mentioned industrial system is used to classify our data set using ANN [2].

In this work, we test the performance of clustering based classification methods on a texture based feature set that has been used in previous studies. Different clustering methods have been applied to the data set and the results are evaluated by using several cluster validation techniques.

There are two main motivation points behind this study:

1) To evaluate the performance of different clustering algorithms for marble classification and to compare these results with ANN results [2]. This information is useful for the applications in which there is not enough number of samples to train a network but still an automatic and objective classification is needed.

2) To determine the best parameter set for a clustering technique that is used for classification of natural stones. To this end, we examine various useful techniques and their effects on determining clustering parameters.

The rest of this paper is organized as follows: Section 2 describes the automated industrial marble inspection system for quality classification. Section 3 introduces our marble slab data set, quality groups and challenges. Section 4 explains the feature extraction strategy. Section 5 gives a brief overview of clustering methods used in this study while Section 6 covers cluster validation methods. Section 7 is composed of simulation results while Section 8 includes our conclusions.

2. Automatic Classification System

The electro-mechanical system in Figure 2, which automatically classifies marble slabs while they are on a conveyor belt and groups them with the help of a control mechanism, has been designed for industrial applications. The developed system is composed of both software and hardware parts. The hardware part is responsible for grouping the marbles mechanically using a conveyor belt, a serial port communication system, pneumatic pistons, a programmable logic controller and their control circuits. The software is responsible for correct determination of quality group for a marble slab and consists of parts that acquire marble images, extract several features, and finally run an ANN algorithm for classification.

For moving conveyor belts, design and implementation of such a system is very challenging and consists of several additional difficulties [3] that necessitate different approaches for overcoming these problems. In marble quality classification tasks, five main steps are implemented:

(i) Recognizing the marble slab on the conveyor belt via a camera and stopping the belt for a short time.

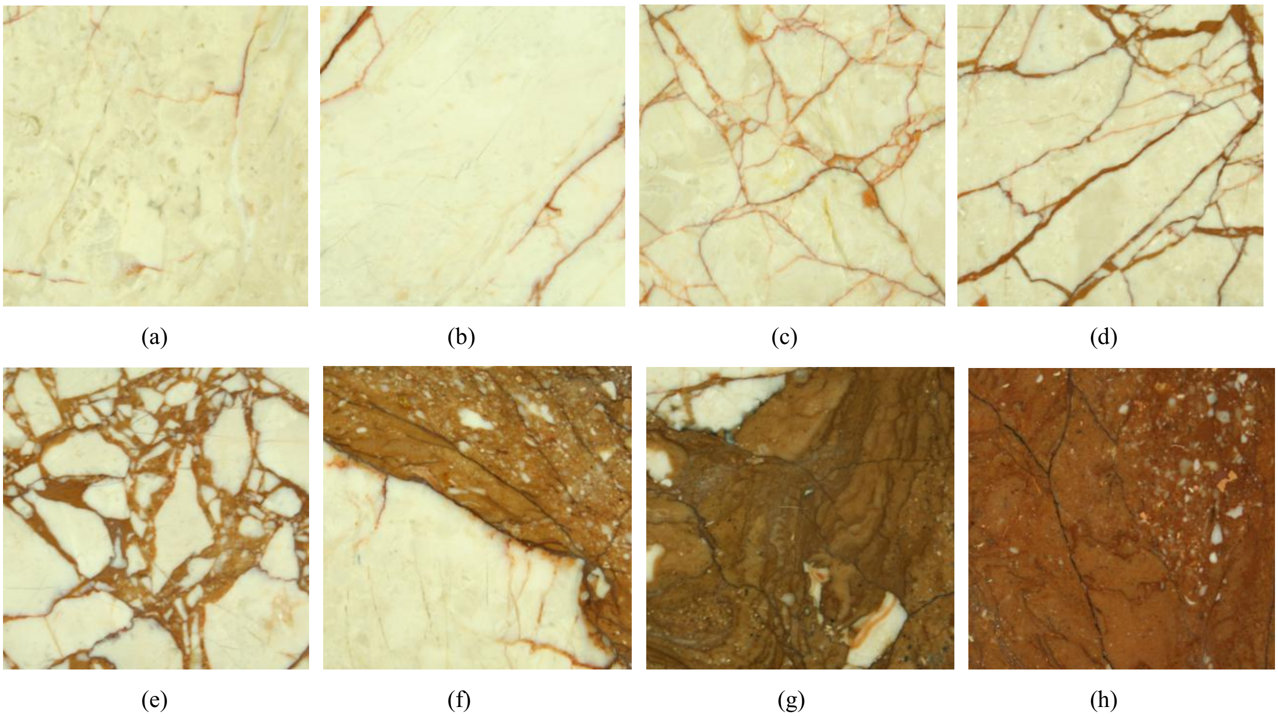

Fig. 1. Typical sample images from each quality group; (a), (b) homogenous limestone, (c), (d) limestone with veins, (e), (f) slabs containing grains (limestone) that are separated by unified cohesive matrix regions, (g), (h) homogenous cohesive matrix.

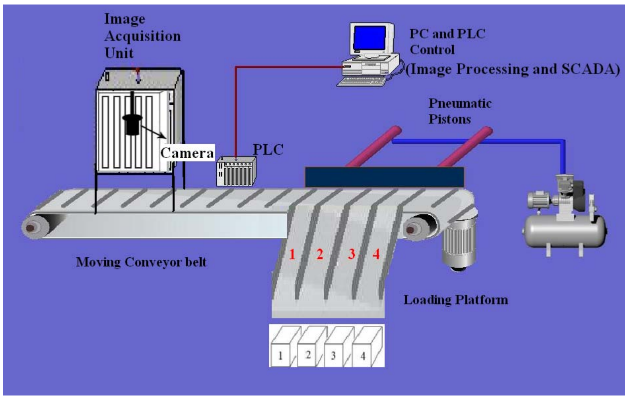

Fig. 2. The scheme of electro-mechanical system for automated industrial marble inspection.

(ii) Acquiring marble image data by taking a picture. The acquisition device is located inside a black box above the conveyor belt to view the items orthographically and to reduce external lighting factors.
(iii) Processing the data to extract several useful features.
(iv) Classification of the marble slab based on the extracted features and using a classifier.
(v) Loading the marble slab to the correct platform depending on the classification done by the classifier. The conveyor belt stops at the correct position and the pneumatic pistons drags the marble slab to the correct loading platform.

The conveyor belt acts as an element that links a production line with the automatic classification system. This allows the possibility of embedding the proposed system into the production line of a marble factory. The short processing time of each marble slab and the obtained high classification percentage rates [2] prove that the system can be integrated to the industry. The proposed system also provides an increase in the quality control standards of the marble slab classification, since marbles are classified with an objective and uniform-through-time criterion.

3. Data Set of Marble Slab Surface Images

The marble slabs are obtained from a mine in Manisa region of Turkey. Ground truth of quality classification is realized by human experts considering color, homogeneity, size, orientation, thickness and distribution of the filled joints and assessing the ratio of limestone grains (beige colored regions in Figure 1) to the cohesive matrix (red-brown colored regions in Figure 1).

Under these criteria, four quality groups have been considered:

1. Homogenous limestone with no or very rare thin joints (beige color) (Figure 1 (a), (b)),
2. Limestone with thin joints (veins) (Figure 1 (c), (d)),
3. Brecciated limestone which is composed of limestone grains of different shape and size cemented with cohesive matrix (Figure 1 (e), (f)). Cohesive matrix is the collection of joints that unify and construct a larger area of red-brown material.
4. Homogenous cohesive matrix with no or very rare limestone (Figure 1 (g), (h)).

Results with ANN show that Groups 1 and 4 can be classified with acceptable performance rates in all systems [2]. However, misclassification ratio is high between the samples of Groups 2 and 3 due to marble slabs having very similar patterns of limestone and cohesive matrix (Figure 1 (d), (e)). This requires advanced classification strategies as proposed in [2], since these groups cannot be separated with high performance.

4. Extraction of Features

For evaluating clustering methods, textural features are extracted by using Sum and Difference Histograms [4] method as briefly explained below. Let us consider a *KxL* image denoted by $\{y_{k,l}\}$, $\{k = 1,2,...,K;\ l = 1,2,...,L\}$ and with quantized gray levels $N_G = 256$. Consider, again, the distance vector $(d1,\ d2) \in D$, which separates two picture elements, $y_1 = y_{k,l}$ and $y_2 = y_{k+d1,l+d2}$, where *D* is the subset of indexes specifying the texture region to be analyzed. Then, the sum, $t_{k,l}$, and difference, $f_{k,l}$, histograms are defined as

$$t_{k,l} = y_{k,l} + y_{k+d1,l+d2}\,, \quad (1)$$

$$f_{k,l} = y_{k,l} - y_{k+d1,l+d2}\,. \quad (2)$$

Here, we define

$$h_t(i) = Card\{(k,l) \in D,\ t_{k,l} = i\}\,, \quad (3)$$

$$h_f(j) = Card\{(k,l) \in D, f_{k,l} = j\}\,, \quad (4)$$

$$N = Card\{D\} = \sum_i h_t(i) = \sum_j h_f(j)\,. \quad (5)$$

Above, $Card\{.\}$ denotes cardinality. Then, the normalized sum and difference histograms are found as

$$P_t(i) = h_t(i)/N, \qquad i = 0,\ ...,\ 2N_G - 2\,, \quad (6)$$

$$P_f(j) = h_f(j)/N,\ \ \mathrm{j} = \text{-}N_G + 1,\ ...,\ N_G\,\text{-}1\,. \quad (7)$$

In [13], the distance metric is chosen to be the 8-neighborhood. Using the obtained SDH vectors, seven statistical features (mean, variance, energy, correlation, entropy, contrast, homogeneity) [4] are computed. For each color channel, these calculations produced a total of 21 (7 features times 3 color channels) features.

5. Clustering Methods for Quality Classification

Clustering [5] is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar based on a distance measure. Clustering is a method of unsupervised learning and a common technique for statistical data analysis used in many fields and applications. The four methods used in this study are described below:

K-Means: K-means algorithm [5] partitions the pixels in the image into *n* clusters by using an iterative procedure. The aim is to minimize the sum, over all clusters, of the within-cluster sums of gray level value-to-cluster centers;

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 .$$

Here, $\|x_i^{(j)} - c_j\|^2$ is a chosen distance (i.e. Euclidean distance) measure between a gray level value $x_i^{(j)}$ and the cluster center c_j. Distance measure is an indicator of the distance of the *n* data points from their respective cluster centers. Also, batch update is implemented where every iteration consists of reassigning feature values to their nearest cluster centers at once. Then, cluster centers are recalculated.

Partitioning Around Medoids (PAM): PAM algorithm uses the dissimilarity matrix of a given data set instead of the Euclidean distance in K-means algorithm [6]. PAM first computes *k* representative objects, called medoids. A medoid can be defined as that object of a cluster, whose average dissimilarity to all the objects in the cluster is minimal. In the classification literature, such representative objects are called centrotypes. After finding the set of medoids, each object of the data set is assigned to the nearest medoid.

Hierarchical Clustering (HC): HC [7] creates a hierarchy of clusters which may be represented in a tree structure called a dendrogram. The root of the tree consists of a single cluster containing all observations, and the leaves correspond to individual observations. Algorithms for HC are generally either agglomerative, in which one starts at the leaves and successively merges clusters together; or divisive, in which one starts at the root and recursively splits the clusters. Any valid metric may be used as a measure of similarity between pairs of observations. Euclidean metric is used in our tests. The choice of which clusters to merge or split is determined by a linkage criterion, which is a function of the pairwise distances between observations. We perform HC via the following steps:

Step 1) Find the similarity or dissimilarity between every pair of objects in the data set. In this step, the distance between objects is calculated.

Step 2) Group the objects into a binary, hierarchical cluster tree. In this step, pairs of objects that are in close proximity are linked using the linkage function. The linkage function uses the distance information generated in 1) to determine the proximity of objects to each other. As objects are paired into binary clusters, the newly formed clusters are grouped into larger clusters until a hierarchical tree is formed.

Step 3) Determine where to cut the hierarchical tree into clusters. In this step, the cluster function to prune branches off the bottom of the hierarchical tree is used, and all the objects below each cut are assigned to a single cluster. This creates a partition of the data. The cluster function can create these clusters by detecting natural groupings in the hierarchical tree or by cutting off the hierarchical tree at an arbitrary point.

Self-Organizing Map (SOM): SOM is a type of ANN that is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples, called a map [8]. SOM uses a neighborhood function to preserve the topological properties of the input space. The goal of learning in the SOM is to cause different parts of the network to respond similarly to certain input patterns. This is partly motivated by how visual, auditory or other sensory information is handled in separate parts of the cerebral cortex in the human brain.

A SOM consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a regular spacing in a hexagonal or rectangular grid. The self-organizing map describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to find the node with the closest weight vector to the vector taken from the data space and to assign the map coordinates of this node to our vector. While it is typical to consider this type of network structure as related to feedforward networks where the nodes are visualized as being attached, this type of architecture is fundamentally different in arrangement and motivation.

6. Validation Methods

Cluster analysis is an important technique in many research areas and as previously mentioned, several clustering algorithms such as K-means, PAM, HC, and SOM are widely used. These clustering methods sometimes give much different or somewhat different solutions. Thus, depending on the data set and application, an appopriate one should be selected. Moreover, the next important step is to adjust the parameters of the selected clustering method and evaluating clustering solutions to determine an optimal solution or cluster structure for the data set, usually the Number of Clusters (NC). This step is called cluster validation which aims to find a solution that gives the best result for a given data set. This is also time consuming because of the many factors to consider (i.e. data preprocessing, similarity metrics, NC, other parameters of clustering algorithms, validity indices, the evaluation of clustering solutions etc.)

For evaluation of clustering solutions, the validity indices are commonly used. There are two kinds of validity indices: external indices and internal indices. An external index is a measure of agreement between two partitions where the first partition is the a priori known clustering structure, and the second results from the clustering procedure. Internal indices are used to measure the goodness of a clustering structure without external information. In our study, we evaluate the clustering results using internal indices and try to determine the optimal NC based on these internal validity indices. Short descriptions of the internal indices used in this paper for estimating NC and evaluating clustering quality are introduced below:

Davies-Bouldin index: A measure of the average similarity between each cluster and its most similar one. Small values correspond to clusters that are compact and have centers that are far away from each other. Therefore, its minimum value determines the optimal NC [9].

Calinski-Harabasz index: The measures of between-cluster isolation and within-cluster coherence. Its maximum value determines the optimal NC [10].

Dunn index: A measure that maximizes the inter-cluster distances while minimizing the intra-cluster distances. Its large values indicate the presence of compact and well-separated clusters, so the NC that maximizes the index is taken as the optimal NC [11].

Silhouette index: A composite index reflecting the compactness and separation of clusters; a larger average Silhouette index indicates a better overall quality of the clustering result. Thus, the optimal NC is the one that gives the largest average Silhouette value [12].

7. Simulations and Results

The simulation results are presented in Fig. 3. In more simple clustering methods (i.e. K-Means, PAM), the index values of cluster validation algorithms produce the best results for NC=4. This is an expected value for NC, since there are in fact four quality groups. However, more complex clustering approaches (i.e. SOM and HC) sometimes give better index values for higher number of clusters, particularly for 9 and 10 clusters. Although there are some disagreements between the results of the validation techniques used in the study, we believe we can conclude that if one uses a simple clustering technique (i.e. K-Means or PAM), simply the number of clusters can be selected equal to the number of quality groups. On the other hand, if one uses a more complex (i.e. SOM) and/or especially a multi-resolution clustering technique (i.e. HC) using more clusters than the number of quality groups may improve the performance.

This conclusion is verified with further simulations. The application of K-Means clustering method results in a best performance of 74 % (average of four quality groups) which is obtained for four clusters. In a similar manner, PAM clustering results in 75 % average correct classification performance, also for four clusters (best result). However, SOM performs its best classification performance for ten clusters with an average performance of 80 %. Between these four clustering techniques, the best performance is obtained using HC with 10 clusters, which performs 83 % of average correct classification.

Considering that the best classification performance in [2] is obtained also with hierarchical classification (i.e. using Hierarchical Radial Basis Function Network) of the data, these results are also in agreement with previous studies.

8. Conclusions

In this study, different clustering methods and several validity indices have been applied to a texture based feature set, which is extracted from a large and diverse set of marble slab images, for quality classification. This is done in order to compare performances of clustering techniques against neural networks. This comparison is important because the clustering techniques might be needed especially when there is not enough number of samples to construct a training set. Although the results are not as high as the ones obtained with ANN basd tecniques, which are around 99 % [2], clustering is still applicable to the natural stone classification applications where a performance around 80 % (more than performance of the human-based classification) is still enough.

Besides performance of the clustering techniques, the influence of the clustering parameters such as number of clusters, validity indices, the evaluation of clustering solutions have been investigated. The results show that, finding the best clustering solution for marble quality classification task depends not only on a validity index but also the appropriate clustering procedure. An obvious result is that the number of clusters can be selected equal to the number of quality groups. However, using different clustering algorithms or different validity indices result in different clustering solutions for this specific clustering task making the process of cluster validation quite complex. One may choose a validity index to estimate an optimal number of clusters, where the optimal clustering solution is found from a series of clustering solutions under different numbers of clusters.

The future studies will cover a wider range of simulations to understand the nature of the clustering process. These simulations will include the observation of more variables such as data preprocessing and other parameters of clustering algorithms. The proposed methodologies will also be applied to different natural stone classification tasks for evaluation.

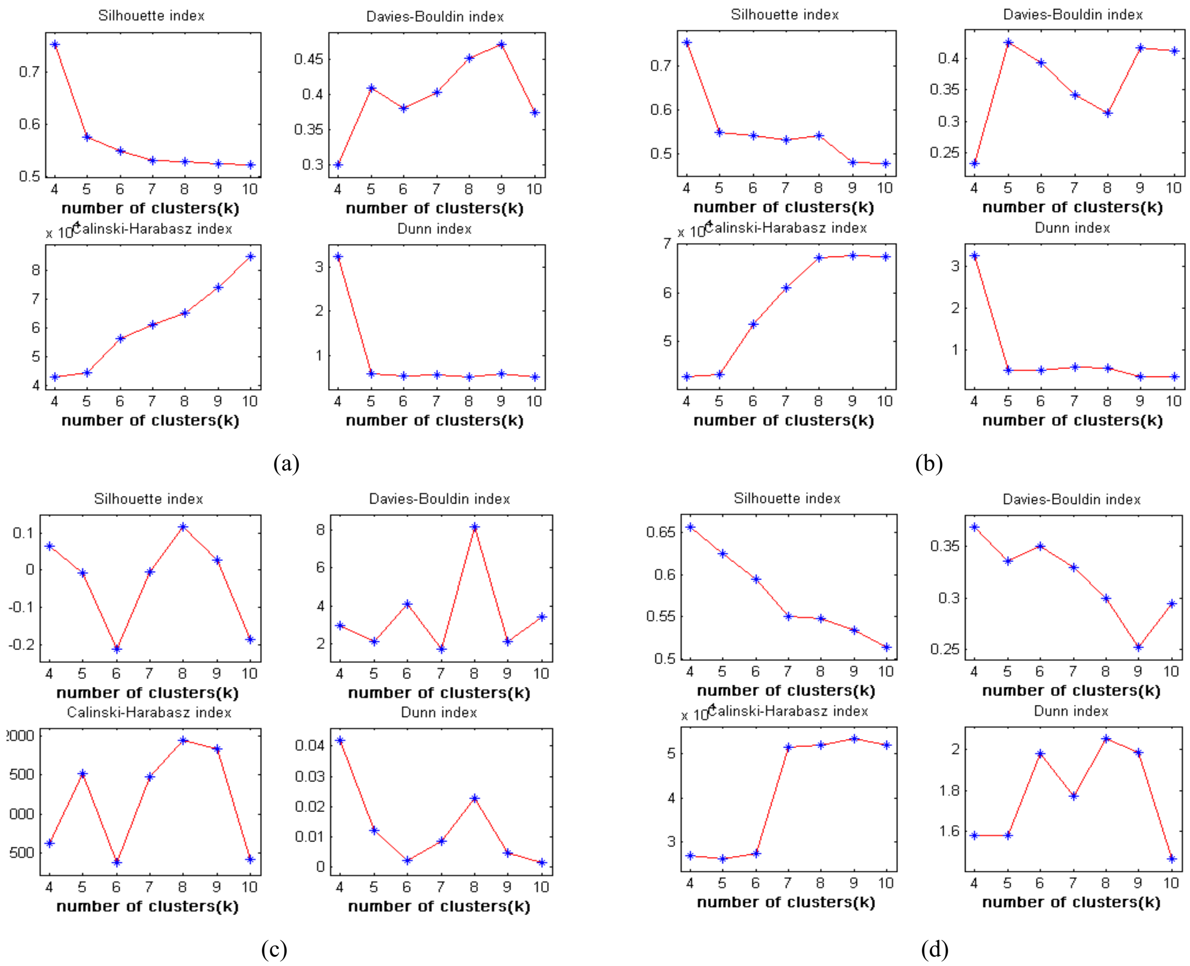

Fig. 3. Clustering validation index graphs for (a) K-Means, (b) PAM, (c) SOM, and (d) HC. The horizontal axes of the graphs represent the number of clusters and the vertical axes show the value of the index of the corresponding validation technique.

9. Acknowledgements

This work was partially supported by TUBITAK - MAG under Grant Number: 104M358.

10. References

[1] A. B. Yavuz, N. Türk and M. Y. Koca, "The use of micritic limestone as building stone. A case study of Akhisar beige marble in western Turkey", Proceedings of Industrial Minerals and Building Stones, pp. 277-281, Turkey, 2003.

[2] M. A. Selver et. al., "Cascaded and hierarchical neural networks for classifying surface images of marble slabs,'' IEEE Trans. on Systems, Man, and Cybernetics - Part C: Applications and Reviews, vol. 39 (4), pp. 426-439, 2009.

[3] T. S. Newman and A. K. Jain, "A survey of automated visual inspection," Comput. Vis. Image Understand., vol. 61, no. 2, pp. 231–262, 1995.

[4] M. Unser, "Sum and difference histograms for texture classification", IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-8, no. 1, pp. 118-125, Jan. 1986.

[5] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proc. 5th Berkeley Symp. Math. Stat. Prob., vol. 281, 1967.

[6] J. C. Gower and P. Legendre, "Metric and Euclidean properties of dissimilarity coefficients," Journal of Classification, vol. 3, pp. 5-48, 1986.

[7] J. H. Ward, "Hierarchical grouping to optimize an objective function," Journal of the American Statistical Association, vol. 58 (301), pp. 236–244, 1963.

[8] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," IEEE Trans Neural Networks, vol. 11 (3), pp. 586-600, 2000.

[9] E. Dimitriadou, S. Dolnicar and A.Weingessel, "An examination of indexes for determining the number of cluster in binary data sets," Psychometrika, vol.67 (1), pp. 137-160, 2002.

[10] S. Dudoit and J. Fridlyand, "A prediction-based resampling method for estimating the number of clusters in a dataset," Genome Biology, 3(7): 0036.1-21, 2002.

[11] M. Halkidi, Y. Batistakis and M. Vazirgiannis, "On clustering validation techniques," Intelligent Information Systems Journal, 17(2-3), pp. 107-145, 2001.

[12] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York, John Wiley & Sons, 1990.