Performance Characterization of Clustering Algorithms for Colour Image Segmentation

Dana Elena Ilea, Paul F. Whelan, Ovidiu Ghita

Vision Systems Group, School of Electronic Engineering

Dublin City University, Dublin 9, Ireland

Abstract-**This paper details the implementation of three traditional clustering techniques (K-Means clustering, Fuzzy C-Means clustering and Adaptive K-Means clustering) that are applied to extract the colour information that is used in the image segmentation process. The aim of this paper is to evaluate the performance of the analysed colour clustering techniques for the extraction of optimal features from colour spaces and investigate which method returns the most consistent results when applied on a large suite of mosaic images**.

***Index terms*: Colour image segmentation, Fuzzy clustering, Competitive Agglomeration, Diffusion based filtering, Image segmentation.**

I. INTRODUCTION

Image segmentation is one of the most important precursors for image processing-based applications and has a decisive impact on the overall performance of the developed system. Robust image segmentation has been the subject for research for many years but the published research indicated that most of the developed image segmentation algorithms have been designed in conjunction with particular applications. The segmentation process consists of separating a digital image into several disjoint regions, whose characteristics such as intensity, colour, and texture are similar. Typically, the goal of segmentation is to locate certain objects of interest in an image. One important area of research is to perform image segmentation that evaluates the similarity of image regions that is defined in terms of colour and texture and this information is used to automatically segment images into semantically meaningful parts [1,2].

In this paper, the colour and texture information are included in a compound mathematical descriptor in order to extract the meaningful regions in an image. In this regard, the texture is analysed using the Local Binary Pattern method and colour can be extracted using clustering methods. The area of colour image analysis is one of the most active topics of research and a large number of colour segmentation techniques have been proposed. Colour segmentation techniques that have been proposed include histogram-based segmentation [3], probabilistic space partitioning and clustering methods [4,5], Markov random field and simulated annealing [6,7].

In this paper we extract the colour information using a statistical analysis of image data using cluster analysis. Its aim is to construct clusters from input data in such a way that the profiles of objects in the same groups are relatively homogenous whereas the profiles of objects in different groups are heterogeneous. Statistical partitioning of data into meaningful clusters is widely used in image analysis since the clustered data has a reduced dimensionality and it is easier to interpret. These algorithms are ideal to extract the colour information since this process will reduce the large number of colour components existing in the initial image into a reduced number of components in the segmented result, which are strongly related to the image objects.

There are many clustering schemes that perform the spatial partition of data using different strategies. In this regard, the ISODATA [8] progressively check inter and intra cluster similarity at each iteration so that it can dynamically split and merge the clusters in an iterative manner. This algorithm is very simplistic and its performance is modest.

In this paper we detail the implementation of three traditional clustering algorithms: K-Means, Fuzzy C-Means and Adaptive K-Means (also called clustering by competitive agglomeration-CA [9] and we evaluate their performance when they are used in a flexible split and merge image segmentation algorithm. To improve the immunity to noise and increase the performance of the colour extraction process, we have used a diffusion based filtering scheme which proved to be an optimal, feature preserving smoothing strategy.

Our aim is to evaluate the performance of three traditional clustering algorithms (K-Means, Fuzzy C-Means and Adaptive K-Means) with the purpose of determining which of them returns optimal results in the colour extraction process.

This paper is organised as follows. Section II describes the three traditional clustering techniques followed by the introduction of the diffusion based filtering method in Section III. Section IV details the image segmentation technique, while Section V presents a number of experimental results. Section VI concludes the paper.

II. CLUSTERING METHODS

A. K-Means Clustering

One of the most popular and widely studied clustering algorithms to separate the input data in the Euclidian space is the K-Means clustering. It is a nonhierarchical technique that follows a simple and easy way to classify a given dataset through a certain number of clusters (we need to make an assumption for parameter *k*) that are known *a priori*. The K-Means algorithm is built using an iterative framework where the elements of the data are exchanged between clusters in order to satisfy the criteria of minimizing the variation within each cluster and maximizing the variation between clusters. When no elements are exchanged between clusters, the process is halted. The four steps of this algorithm are briefly described below:

Steps of the K-Means clustering algorithm:

1. Initialization – define the number of clusters and randomly select the position of the centers for each cluster or directly generate k seed points as cluster centers.
2. Assign each data point to the nearest cluster center.
3. Calculate the new cluster centers for clusters receiving new data points and for clusters losing data points.
4. Repeat the steps 2 and 3 until a convergence criterion is met (when there is no exchange of data points between the k clusters).

The aim of the K-Means is the minimization of an objective function:

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

$\left\| x_i^{(j)} - c_j \right\|^2$ is the distance measure (usually Euclidian metric) between a data point $x_i^{(j)}$ and the cluster center c_j (this is an indicator of the distance of the n data points from the cluster centers).

There are situations when the K-Means algorithm doesn't find the optimal solution corresponding to the global objective function J and in addition is sensitive to the initialisation process that selects the initial cluster centers that are usually randomly picked from input data.

The main advantages of this algorithm are its simplicity and low computational cost, which allows it to run efficiently on large datasets. The main drawback is the fact that it does not systematically yield the same result each time the algorithm is executed and the resulting clusters depend on the initial assignments. The K-Means algorithm maximizes inter-cluster (or minimizes intra-cluster) variance, but does not ensure that the algorithm will not converge to local minima due to an improper starting condition (initialization of the cluster centers).

B. Fuzzy C-Means Clustering

Fuzzy C-Means is a clustering method that was first developed by Dunn in 1973 [10] and improved by Bezdek in 1981 [11]. The fuzzy set theory allows an element of the data to belong to a cluster with a degree of membership that has a value in the interval [0,1]. For K-Means clustering the degree of membership for a pattern in a particular cluster is 1 if the pattern belongs to the cluster or 0 if it doesn't. In fuzzy set theory, a pattern can belong to 2 or more clusters simultaneously where the membership grades determine the degree to which the pattern belongs to these clusters. Like K-Means, Fuzzy C-Means tends to minimize the following objective function:

$$J = \sum_{j=1}^{C}\sum_{i=1}^{n}(u_{ij})^m\left\| x_i^{(j)} - c_j \right\|^2 \tag{2}$$

where u_{ij} are the membership values that form the membership matrix U (represents the degree of membership to the C clusters and it has a value in the range of [0 1] for each feature vector x_i to the fuzzy cluster c_j. The parameter m is the called the fuzzifier factor and determines the level of cluster fuzziness. A large value for m results in smaller membership u_{ij} and hence fuzzier clusters.

Steps of the Fuzzy C-Means Clustering Algorithm

1. Consider a set of n data points (vectors) to be clustered.
2. Assume the number of clusters C is known: $C \in [2,n)$.
3. Choose an appropriate level of cluster fuzziness $m \in R$, $m > 1$.
4. Initialize the ($n \times c$) size membership matrix U to random values such that:
$$u_{ij} \in [0\ 1] \text{ and } \sum_{j=1}^{C} u_{ij} = 1. \tag{3}$$
5. Calculate the cluster centers c_j using
$$c_j = \frac{\sum_{i=1}^{n}(u_{ij})^m x_i}{\sum_{i=1}^{n}(u_{ij})^m}, \quad \text{for } j = 1 \ldots C. \tag{4}$$
6. Calculate the distance measures $d_{ij} = \left\| x_i^{(j)} - c_j \right\|$ for all clusters $j = 1 \ldots C$ and data points $i = 1 \ldots n$.
7. Update the fuzzy membership matrix U according to d_{ij}. If $d_{ij} > 0$ then $u_{ij} = \left[\sum_{k=1}^{C}\left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}} \right]^{-1}$. If $d_{ij} = 0$ then the data point x_j coincides with the cluster center c_j and so full membership can be set $u_{ij} = 1$.
8. Loop from step 5 until the change in U is less than a given tolerance.

The objective of the Fuzzy C-Means algorithm is the minimization of the intra-cluster variability. Fuzzy C-Means shares the same problems as *K*-Means algorithm, namely the initial selection of the cluster centers and the initial choice of weights u_{ij}. The advantages consists in the idea that fuzzy clustering allows overlapping clusters with partial membership of individual clusters and we can control the level of cluster fuzziness. Because there is no total comitment of a given point to a given cluster, fuzzy clustering algorithms require more memory and are computationaly expensive when applied to large datasets.

C. Adaptive K-means Clustering

The Adaptive K-Means scheme is a competitive agglomeration (CA) clustering algorithm that has been developed by Frigui and Krishnapuram [9]. The objective of the CA clustering algorithm is the minimization of an objective function that incorporates the advantages of both hierarchical and contextual clustering. The CA algorithm produces a sequence of partitions where the number of clusters decreases at each iteration. The initial partition of data has an over specified number of clusters, and the final partition converges to an "optimal" number of clusters. The update equation in the CA algorithm creates an environment in which clusters compete for feature points and only clusters with large cardinalities survive [9].

As it was mentioned above, the aim of the CA algorithm is the minimization of an objective function that is illustrated in the following equation:

$$J = \sum_{j=1}^{C}\sum_{i=1}^{n}(u_{ji})^m \left\| x_i^{(j)} - c_j \right\|^2 - \alpha \sum_{j=1}^{C}\left[\sum_{i=1}^{n} u_{ji}\right]^2 ,$$

$$\text{and} \quad \sum_{j=1}^{C} u_{ji} = 1 \ \text{ for } i \in \{1 \ldots \text{n}\}. \tag{5}$$

It should be noticed the two components of the objective function. The first term is similar to the Fuzzy C-Means objective function. The second component is the sum of squares of the cardinalities of the clusters that allows the algorithm to control the number of clusters and reduce them progressively.

Combining these components and making a proper choice for parameter α, the final partition will minimize the sum of intra-cluster distances, while partitioning the data set into the smallest possible number of clusters.

Steps of the CA Algorithm:

1. Fix the maximum number of clusters $c = c_{\max}$;
2. Initialize iteration counter $k = 0$.
3. Initialize the fuzzy C membership matrix *U*.
4. Compute the initial cardinalities n_j for $1 \le j \le c$, using equation $n_s = \sum_{i=1}^{n} u_{si}$ (the cardinality of a cluster *s*).
5. Loop in the following steps until the prototype parameters stabilize:
   a. Compute $d^2(x_i, c_j)$ for $1 \le i \le n,\ 1 \le j \le c$;
   b. Update $\alpha(k)$ by using equation

$$\alpha(k) = \eta_0 \exp(-k/\tau) \frac{\sum_{j=1}^{C}\sum_{i=1}^{n}(u_{ji})^2 d^2(x_i, c_j)}{\sum_{j=1}^{C}\left[\sum_{i=1}^{n}(u_{ji})\right]^2} \tag{6}$$

where: $\eta_0 \exp(-k/\tau)$ is the exponential decay factor, *k* is the number of iterations and τ is the time constant.

   c. Update the partition matrix by using the equation

$$u_{st} = u_{st}^{FCM} + u_{st}^{Bias};$$

   d. Compute the cardinality n_j for $1 \le j \le c$ using the equation: $n_s = \sum_{i=1}^{n} u_{si}$.
   e. If $(n_j < \varepsilon_1)$ discard cluster c_j;
   f. Update the number of clusters *c*;
   g. Update the prototype parameters *k=k+1*;
6. Until (prototype parameters stabilize).

The main advantage of the hierarchical procedures is that the clustering is not influenced by the starting conditions and the number of clusters doesn't need to be mentioned *a priori*. The main disadvantage of hierarchical clustering procedures is that they partition the data statistically and no information about the global shape or size of clusters is used in the clustering process. Also, this type of clustering has a static character because the points that are committed to one cluster in the early stages cannot move to another cluster since the sequence of partitions is nested.

III. DIFFUSION-BASED FILTERING TECHNIQUE

Diffusion based filtering was originally developed by Perona and Malik with the purpose of implementing an optimal, feature preserving smoothing strategy [12]. In their paper, smoothing is formulated as a diffusive process and is performed within the image regions and suppressed at the regions boundaries. The standard linear smoothing filtering based on local averaging or Gaussian weighted spatial operators reduces the level of noise but this advantage is obtained at the expense of poor feature preservation (i.e. narrow details in the image). For this implementation we have applied a smoothing filtering scheme based on anisotropic diffusion where smoothing is performed at intra regions and suppressed at region boundaries [13]. This nonlinear smoothing procedure can be defined in terms of the derivative of the flux function:

$$u_t = div(D(|\nabla u|)\nabla u) \tag{7}$$

where u is the input data, D represents the diffusion function and t indicates the iteration step. The smoothing strategy described in equation (7) can be implemented using an iterative formulation as follows:

$$I_{x,y}^{t+1} = I_{x,y}^{t} + \lambda \sum_{j=1}^{4} [D(\nabla_j I)\nabla_j I] \tag{8}$$

$$D(\nabla I) = e^{-\left(\frac{\nabla I}{d}\right)^2} \in (0,1] \tag{9}$$

where $\nabla_j I$ is the gradient operator defined in a 4-connected neighborhood, λ is the contrast operator that is set in the range $0< \lambda<0.16$ and d is the diffusion parameter that controls the smoothing level. It should be noted that in cases where the gradient has high values the value for $D(\nabla I)\rightarrow 0$ and the smoothing process is halted.

IV. IMAGE SEGMENTATION

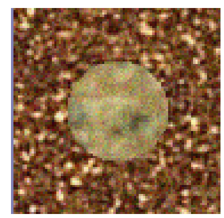
The image segmentation method is based on a split and merge algorithm [14]. The first step of the algorithm recursively splits the image hierarchically into four sub-blocks using the texture information extracted using the LBP/C method. The splitting decision evaluates the uniformity factor of the region under analysis that is sampled using the Kolmogorov-Smirnov metric. In this regard, the pairwise similarity values of the four sub-blocks are calculated and the ratio between the highest and lowest similarity values are compared with a threshold value (split threshold). The region is split if this ratio is higher than the split threshold. This process continues until the uniformity level set by the split threshold is upheld or the block size is smaller than a predefined threshold value (for this implementation the smallest block size has been set to 16×16). During the splitting process two distributions are computed, the LBP/C distribution and the distribution of labels contained in the clustered data.

The second step applies an agglomerative merging procedure on the image resulting after splitting in order to join the adjacent regions that have similar characteristics. This procedure calculates the merging importance between all adjacent regions in the split image and the adjacent regions with the smallest merging importance value are merged. The resulting image has a blocky structure since the regions resulting from splitting process are rectangular. To compensate for this issue the last step of the algorithm applies a pixelwise procedure that exchange the pixels situated at the boundaries between various regions using the colour information sampled by the clustered image calculated with one of the methods detailed in Section II. For more details about the image segmentation algorithm the reader can refer to the following paper [14].

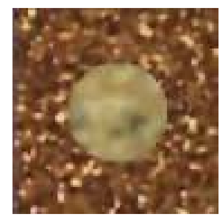
V. EXPERIMENTS AND RESULTS

The clustering algorithms detailed in the Section II have been applied on a large number of images in order to evaluate their performance in regard to the extraction of the colour information. In our experiments we constructed mosaic images using the VisTex database.

The input images are filtered with the diffusion-based method before clustering to reduce the level of noise and improve the colour homogeneity. The parameter for diffusion is set to 10 and the algorithm is iteratively applied 20 times. To show the efficiency of this algorithm the input images were corrupted with Gaussian noise (standard deviation 30 greyscale values) and experimental results are shown in Fig 1.



M1 (a)



M1 (b)

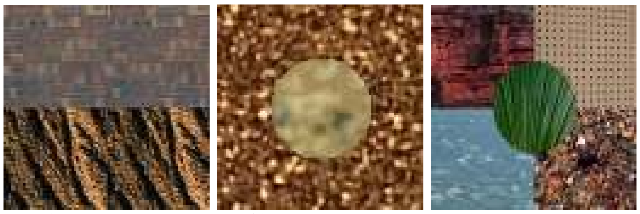
M2 (a)

M2 (b)

Fig. 1. Diffusion based filtering technique applied on mosaic images. (a) Input image corrupted with Gaussian noise and (b) Filtered image.

To evaluate the influence of colour information in the segmentation process we tested the three standard clustering strategies previously discussed. Based on experiments made on 30 different mosaic images, it has been noticed that the outputs of the segmentation procedure is greatly influenced by the appropriate selection of merge and threshold parameters and also by the number of clusters required by each clustering scheme. In this paper we focus on the evaluation of the clustering schemes that are applied to the extraction of colour information.

The experimental data indicated that the performance of the Fuzzy C-Means is closely related to that of the K-Means algorithm and the results indicated the algorithm performed best when the fuzziness parameter has been set close to 1, which translates in an operation close to a crisp clustering scheme similar to the K-Means. As indicated in Section II, the adaptive clustering scheme is derived from Fuzzy C-Means where is evaluated an additional cardinality factor that controls to which extent the clusters are merged together during the clustering process. The performance of the adaptive clustering scheme is generally superior to the performance of the K-Means and Fuzzy C-Means but the cardinality factor is a parameter difficult to be controlled for a large number of images.



a) Image "M1" b) Image "M2" c) Image M3"

Fig.2. Three mosaic images selected to illustrate the segmentation error for each clustering method where the number of clusters is a parameter.

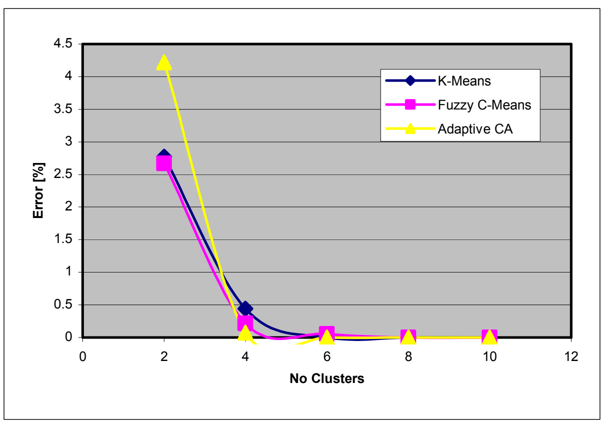


Fig. 3. The segmentation error when the K-Means, Fuzzy C-Means and Adaptive CA clustering algorithms are used to extract the colour information from Image "M1".

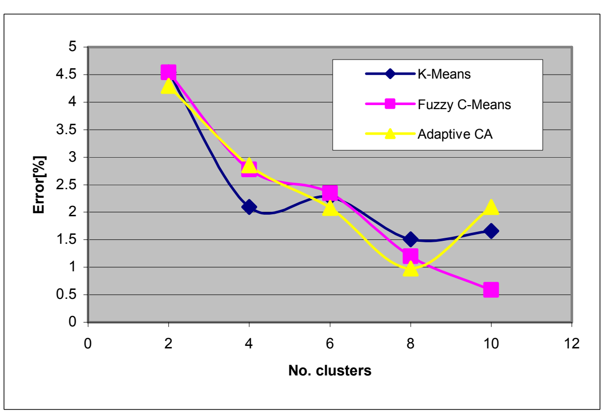


Fig. 4. The segmentation error when the K-Means, Fuzzy C-Means and Adaptive CA clustering algorithms are used to extract the colour information from Image "M2"

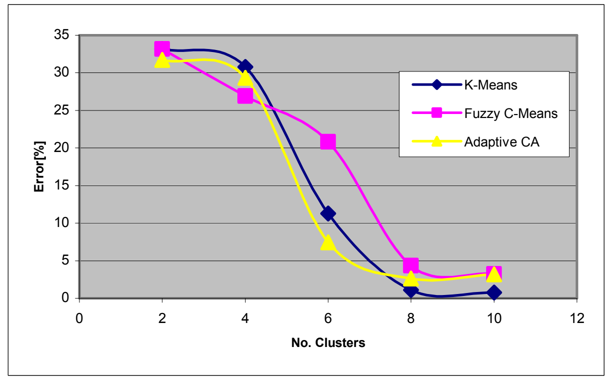


Fig. 5. The segmentation error when the K-Means, Fuzzy C-Means and Adaptive CA clustering algorithms are used to extract the colour information from Image "M3"

Figs. 3 to 5 show the segmentation error when the image segmentation algorithm has been applied to three representative mosaic images depicted in Fig. 2. For these tests the split and merge parameters of the segmentation are maintained constant and only the number of clusters is varied. The error was calculated by comparing the results to the manually segmented ground truth.

Fig. 6 depicts the results obtained after the pixelwise stage of the segmentation process is applied.

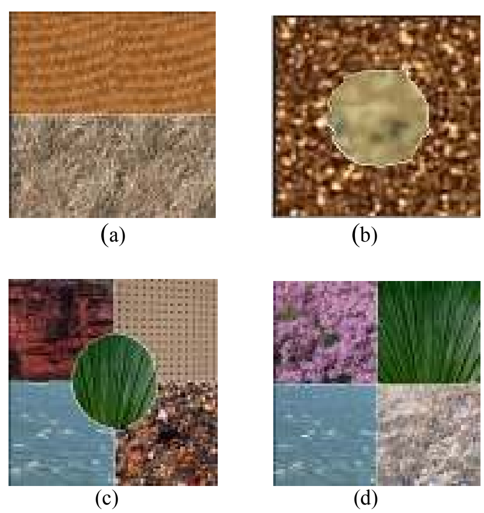


Fig. 6. Colour -texture segmentation of mosaic images.

The graphs depicted in Figs. 3 to 5 indicate that best results are obtained when the number of clusters is selected in the range 6 to 10. To fully evaluate the influence of this parameter on the overall performance of colour clustering algorithms in addition to the three standard clustering methods we have developed an unsupervised K-Means algorithm that is able to determine the optimal number of clusters by evaluating the intra cluster variability (the algorithm adaptively merge the clusters with similar characteristics). The results illustrated in Table I and Table II are obtained when the K-Means algorithm is applied when the number of clusters is selected to 8 and the results when the algorithm is applied in the unsupervised form (the number of clusters are determined adaptively for each image). We can notice the improved performance when the K-Means algorithm is applied in the unsupervised form.

TABLE I
K-MEANS CLUSTERING RESULTS

Test Data	Method Used	No. of Clusters	Average Error [%]	Standard Deviation
30 Mosaic Images	K-Means Clustering (using LTI)	8	2.87	4.76

TABLE II
UNSUPERVISED K-MEANS CLUSTERING RESULTS

Test Data	Method Used	No. of Clusters	Average Error [%]	Standard Deviation
30 Mosaic Images	Unsupervised K-Means Clustering	Automatically detected	2.40	3.03

VI. CONCLUSIONS

In this paper we detailed the implementation of three traditional clustering algorithms (K-Means clustering, Fuzzy C-Means clustering and Adaptive K-Means clustering). To fully compare the efficiency of these clustering techniques we have evaluated the segmentation error when the colour information is used along with texture as a feature in a split and merge image segmentation algorithm. In order to evaluate the performance of the clustering algorithm in the extraction of colour information the influence of texture in the segmentation process has been weakened. These algorithms have been applied on a suite of 30 mosaic images and a large number of tests have been performed to identify their performance. The experimental data indicate that neither method clearly emerge as a winner and for all clustering methods evaluated in this paper the selection of optimal parameters proved to be crucial in obtaining appropriate image segmentation. This conclusion is supported by the experimental data that indicate that far better results are obtained when the clustering algorithms are designed to select adaptively the optimal set of parameters.

REFERENCES

[1] F.Y. Shih and S. Cheng, *Automatic seeded region growing for colour image segmentation*. Science Direct, May 2005

[2] H.D. Cheng, X. Jiang, Y. Sun, J. Wang, “Colour image segmentation: advances and prospects”, *Pattern Recognition* 34(12), pp. 2259-2281, 2001.

[3] L. Shafarenko, M. Petrou, and J. Kittler, “Histogram-based segmentation in a perceptually uniform colour space”, *IEEE Trans. on Image Processing*, vol. 7, no. 9, pp. 1354-1358, 1998.

[4] D. Comaniciu and P. Meer, “Mean Shift: A robust approach toward feature space analysis”, *IEEE Trans. Pattern Analysis Machine Intell.,* vol. 24, no. 5, pp. 603-619, 2002.

[5] H.T. Zöller, and, J. M. Buhmann, “Parametric distributional clustering for image segmentation”, *7th European Conference on Computer Vision (ECCV)*, Copenhagen, Denmark pp. 577-591, 2002.

[6] S. Lakshmanan and H. Derin, “Simultaneous parameter estimation and segmentation of Gibbs Random Fields using simulated annealing”, *IEEE Trans. Pattern Analysis Machine Intell.*, vol. 11, no. 8, pp. 799 – 813, 1989.

[7] J. Mukherjee, “MRF Clustering for segmentation of colour images”, *Pattern Recognition Letters*, vol. 23, no. 8, pp. 917-929, 2002.

[8] A.K. Jain and R.C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ, 1988.

[9] H. Frigui and R. Krishnapuram. “Clustering by Competitive Agglomeration”, *Pattern Recognition*, vol. 30, no. 7, pp. 1109-1119, 1997.

[10] J.C. Bezdek, and J.C. Dunn., “Optimal fuzzy partitions: a heuristic for estimating the parameters in a mixture of normal distributions”, *IEEE Trans. Comput.*, August 1975.

[11] Bezdek, J. C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[12] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion”, *IEEE Trans. Pattern Analysis Machine Intell*, vol. 12, no. 7, pp. 629-639, 1990.

[13] O. Ghita, K. Robinson, M. Lynch, P.F. Whelan, “MRI diffusion-based filtering: a note on performance characterization” *Computerized Medical Imaging and Graphics*, vol. 29, pp. 267-277, 2005.

[14] P. Nammalwar, O. Ghita, P.F. Whelan, “Integration of feature distributions for colour texture segmentation” ICPR 2004. *Proc. of the 17th International Conference on Pattern Recognition (ICPR 2004)*, vol. 1, pp. 716 – 719, 23-26 Aug. 2004.