# Fast Color Image Segmentation using Fuzzy Clustering

*Patrick Lambert*
*Universite de Savoie, France*

*Horia Grecu*
*Universitatea Politehnica Bucuresti, Romania*

## Abstract

In this paper we focus on the problem of image segmentation by color clustering. We present a robust agglomerating clustering algorithm based on cluster validity criteria derived from fuzzy partitions. The result is a simplified segmentation having a small number of large regions. The interest of the proposed method is that it requires a single parameter and that the computational complexity is very low.

## 1 Context

Segmentation is a very important task which is difficult because it both depends on the image type and on the aim of the analysis.[1-4] Clustering is a common step to get a segmentation. In the data space, clusters are regarded as regions of high density which are separated by regions of low density.

Most popular clustering algorithms are suffering of two major drawbacks. First, the number of clusters is predefined, which makes them inadequate for batch processing of huge image databases. Secondly, the clusters are represented by their centroid and built using an Euclidean distance therefore inducing generally an hyperspheric cluster shape, which makes them unable to capture the real structure of the data. This is especially true in the case of color clustering where clusters are arbitrarily shaped.

The algorithm we propose belongs to the class of hierarchical agglomerative algorithms and is based on fuzzy estimators for *isolation* and *compactness*. Basically, it consists in starting with an initial clustering and iteratively merging clusters in larger ones at each step. The key idea is that merging is performed pairwise only if the *isolation* between two clusters is comparable with their *compactness*. Using a small initial number of clusters, the result of the segmentation is a simplified segmentation having a small number of large regions.

## 2 Agglomerative Clustering

### 2.1 Initialization

The initial clusters are computed using the Fuzzy C-Mean (FCM) algorithm[1,5] well-known for its performances in terms of robustness to outliers and speed, but any other clustering technique can be used in this step as long as it produces a sufficiently large number of clusters with respect to the actual number.

Let $X = \{x_1, x_2, \dots x_n\}$ be the set of vectors to be clustered and $v_1, \dots, v_c$ the centers of the classification, where $c$ denotes the number of classes.

The basic idea of the FCM clustering is to assign to each vector $x_{k\ ,\ k\ =\ 1\ \dots\ n}$, a membership degree, $u_{ik}$, to each class $C_i$ centred in $v_i$. The algorithm minimizes a certain intra-class error by iteratively computing the membership degrees and the class centers. The clusters produced by this technique have generally an hyperspherical shape.

### 2.2 Agglomeration is Necessary

Since the shape of the real clusters is arbitrary and is not necessarilly hyperspherical, the classes given by the FCM clustering are no longer a valid representation. So, we represent each real cluster by a set of hyperspherical clusters (called sub-clusters in the following) given by the FCM clustering. The idea is that arbitrary shaped clusters can be approximated by a set of $p$ hyperspheres covering the initial shape.

Let C be an arbitrarily shaped cluster:

$$C = C_1 \cup C_2 \cup \dots \cup C_p \qquad \text{(Eq. 1)}$$

The difficulty is to define a relevant method to merge the different hyperspherical sub-clusters to get a valid cluster. This is performed by defining and comparing the isolation the compactness of the different clusters.

### 2.3 Definition of Isolation and Compactness

A valid cluster has to be both *compact* and *isolated* from others valid clusters. This requires the definition of *Isolation* and *Compactness* of a cluster.

The **isolation** estimator is computed from the fuzzy partition matrix $U = [u_{ik}]$, $1 \le i \le c$, $1 \le k \le n$. Each column $u_i$ can be regarded as a fuzzy cluster. Let $u_i$ and $u_j$ be two fuzzy clusters. The fuzzy intersection and union of the two clusters are defined by:

$$\overset{i,j}{E\cap} = u_i \cap u_j = \{\min(u_{i,1}, u_{j1}), \dots, \min(u_{in}, u_{jn})\} \qquad \text{(Eq. 2)}$$

$$\overset{i,j}{E\cap} = u_i \cap u_j = \{\max(u_{i,1}, u_{j1}), \dots, \max(u_{in}, u_{jn})\} \qquad \text{(Eq. 3)}$$

The intersection of two fuzzy sets is a measure of how much the two sets overlap and therefore can be used as an estimator for how isolated two clusters are with respect to each other.[7]

We introduce the following criterion for measuring the mutual isolation of two fuzzy clusters $u_i$ and $u_j$:

$$S_{ij} = \frac{Card_{Fuzzy}(u_i \cap u_j)}{Card_{Fuzzy}(u_i \cup u_j)} = \frac{\sum_{k=1}^{n} \min(u_{ik}, u_{jk})}{\sum_{k=1}^{n} \max(u_{ik}, u_{jk})} \qquad \text{(Eq. 4)}$$

For $S_{ij}$ close to 0, the clusters $u_i$ and $u_j$ are well isolated, while for $S_{ij}$ close to 1 the clusters are completely overlapping.

As we mentioned earlier, we can define **compactness** in terms of isolation between building clusters.

### 2.4 Merging Process

The merging process is based on binary merging decision for each pair of clusters.

For two clusters $C_i$, $C_j$, at any iteration $t$, we are comparing the mutual isolation $I_{ij}$ with the average compactness of the clusters:

$$T_{ij}^t = \frac{Card_{Fuzzy}(C_i^t) \times CP(C_i^t) + Card_{Fuzzy}(C_j^t) \times CP(C_j^t)}{Card_{Fuzzy}(C_i^t) + Card_{Fuzzy}(C_j^t)} \qquad \text{(Eq. 6)}$$

As stated before, the merging decision is made only if the isolation is small with respect to the average compactness.

$$T_{ij}^t = \alpha \cdot S_{ij}^t \qquad \text{(Eq. 7)}$$

where is the only parameter that controls the clustering coarseness.

At each iteration step, all possible merging are executed and the fuzzy partition matrix recomputed using a distance between a vector $x$ and the cluster $C$ given by (Eq. 8):

$$\begin{aligned} D(x,C) &= \min(D(x,C_1), D(x,C_2), \ldots, D(x,C_p)) \\ &= \min\left(\|x - v_1\|, \|x - v_2\|, \ldots \|x - v_p\|\right) \end{aligned} \qquad \text{(Eq. 8)}$$

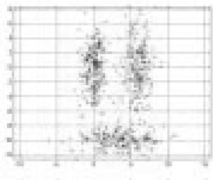
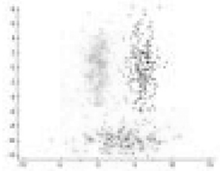
and where ‖ . ‖ is a distance function. When no more merging is allowed, the process stops.

## 4 Results

In our experiments we used the chrominance components of the La*b* space. The luminance was not taken into account in order to provide robustness to illumination changes.

We first present a result on synthetic generated data, which is proving the ability of the algorithm to detect the correct number and shape of the clusters.

In the following example (in the case of a two dimensional space), we can see that the algorithm gives a successful classification in two classes (fig 1-b) despite the non-spherical shape of the initial data (fig 1-a). The initial clustering uses 16 clusters.

a

b


*a: initial data* *b: classification result*

*Figure 1. Classification result on synthetic data*

The second result is obtaibed on natural images. The image segmentation was performed by color clustering in the a*b* space. We are therefore taking into account only pixels with luminance between a minimal and a maximal luminance threshold.

Figures 2 gives some segmentation results on rather complex natural images. We can note that the method provides a good segmentation with a small number of relevant regions, even in the case of textured background.

## 5 Conclusion

In this paper, we proposed a non-parametric unsupervised clustering algorithm based on cluster validity criteria derived from fuzzy partitions. We presented application of the algorithm to color image. The result is a simplified segmentation which can be used in any problem where the required segmentation may be an approximate one.[6]

## References


1. J. Bezdeck, “Pattern recognition with fuzzy objective functions algorithms”, Plenum Press, New York, 1982.
2. A. N. Zahid. O. Abouelala, M. Limouri, “Unsupervised fuzzy clustering”, Pattern Recognition Letters, 20 (1999), pp. 123-129.
3. T. Uchiyama & M. A. Arbib, “Color Image Segmentation using competitive learning”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. PAMI-16, n° 12, 1994, pp. 1197-1206.
4. P. Scheunders, “A genetic c-means clustering algo-rithm applied to color image quantization,” Pattern Recognition, vol. 30, no. 6, pp. 859–866, 1997.
5. Y.W. Lim & S.U. Lee, “On the color image segmentation algorithm based on the thresholding and the fuzzy c-Means techniques”, Pattern Recognition, vol. 23, no 9, 1990, pp. 935-952.
6. H. Grecu & P. Lambert, “Image Retrieval by Color Signatures”, to be published in Proc. of the International Workshop on Content-Based Multimedia Indexing (CBMI’01), Brescia, Italy, Sept. 2001.


*a: initial image* *b: segmented image* *a: initial image;* *b: segmented image*

*Figure 2.*