



**HAL**  
open science

# On Selecting an Optimal Number of Clusters for Color Image Segmentation

Hoel Le Capitaine, Carl Frelicot

► **To cite this version:**

Hoel Le Capitaine, Carl Frelicot. On Selecting an Optimal Number of Clusters for Color Image Segmentation. Proceedings of the 2010 20th International Conference on Pattern Recognition, 2010, Istanbul, Turkey. pp.3388–3391. hal-00565945

**HAL Id: hal-00565945**

**<https://hal.science/hal-00565945>**

Submitted on 15 Feb 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On selecting an optimal number of clusters for color image segmentation

Hoel Le Capitaine

MIA Laboratory, University of La Rochelle

La Rochelle, France

Email: hoel.le\_capitaine@univ-lr.fr

Carl Frélicot

MIA Laboratory, University of La Rochelle

La Rochelle, France

Email: carl.frelicot@univ-lr.fr

**Abstract**—This paper addresses the problem of region-based color image segmentation using a fuzzy clustering algorithm, e.g. a spatial version of fuzzy c-means, in order to partition the image into clusters corresponding to homogeneous regions. We propose to determine the optimal number of clusters, and so the number of regions, by using a new cluster validity index computed on fuzzy partitions. Experimental results and comparison with other existing methods show the validity and the efficiency of the proposed method.

**Keywords**-cluster validity, clustering methods, color image segmentation, overlap and separation measures

## I. INTRODUCTION

Region-based image segmentation consists in partitioning an image into non-intersecting regions such that pixels within a region are homogeneous and pixels from adjacent regions are not. Fuzzy clustering algorithms have become popular, e.g. fuzzy c-means (FCM) [1]. Like any unsupervised algorithm, it requires the number of clusters to be set by the user. This makes FCM unsuited for segmenting image databases unless this number is automatically determined. Cluster validity indexes (CVIs) come to this end and once an optimal number of clusters for a dataset is obtained, the corresponding fuzzy partition is selected as the optimal one [2], [3], [4]. Usual FCM and its spatial version [5] we use are presented in section II. In section III, we recall basic definitions on fuzzy aggregation operators [6] and present a CVI to be used in an automatic segmentation process. Experimental results are given in section IV, comparison to other unsupervised segmentation algorithms is provided. We finally conclude and draw some perspectives in section V.

## II. FUZZY CLUSTERING AND IMAGE SEGMENTATION

Clustering is an instance of unsupervised classification which aims at finding a structure of groups in a set of  $n$   $p$ -dimensional points  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ . In this framework, the label vectors  $\mathbf{u}_k = u(\mathbf{x}_k)$  do not exist and clustering algorithms can be used to obtain them from  $X$ . For instance, the *fuzzy c-means* (FCM) algorithm [1] partitions  $X$  into  $c > 1$  clusters by minimizing the following objective function

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (1)$$

where  $u_{ik}$  is the membership degree of  $\mathbf{x}_k$  to the  $i^{\text{th}}$  cluster represented by its centroid  $\mathbf{v}_i \in \mathbb{R}^p$ . Centroids are gathered into a  $(c \times p)$  matrix  $V = [\mathbf{v}_1, \dots, \mathbf{v}_c]$ . Degrees  $u_{ik}$  are subject to  $\sum_{i=1}^c u_{ik} = 1$  for all  $\mathbf{x}_k$  in  $X$  and to  $0 < \sum_{k=1}^n u_{ik} < n$  ( $\forall i = 1, \dots, c$ ). In addition, they are elements of the fuzzy  $c$ -partition matrix  $U$  ( $c \times n$ ). The so-called *fuzzifier*  $m > 1$  is a weighting exponent which makes the resulting partition more or less fuzzy [4]. The higher  $m$  is, the softer the clusters' boundaries are. Minimization of  $J_m(U, V)$  is obtained by iteratively and alternatively updating  $(U, V)$ . Application to image segmentation consists in taking  $X$  as the entire set of pixels  $\mathbf{x}_k$  in a color space, say  $\mathbb{R}^3$ , and obtain  $r$  regions given by  $(U, V)$ , where  $r \geq c$  since several non adjacent regions in the image can have the same centroid in the color space. Unfortunately, FCM fails when noisy pixels are present so a filtering process is required. An alternative is to add spatial constraints in order to smooth heterogeneous pixel values effect within a region, e.g. the simple sFCM algorithm [5] where the degrees of neighboring pixels are used to modify  $U$  during the updating step as follows :

$$u_{ik} = \frac{u_{ik}^p h_{ik}^q}{\sum_{j=1}^c u_{jk}^p h_{jk}^q} \quad (2)$$

where  $h$  is a spatial function and  $(p, q)$  are parameters which control the relative importance of  $u_{jk}$  and  $h_{jk}$  so that the smoothing effect due to the neighborhood increases with  $q$ . The general form of  $h$  is given by  $h_{ik} = \sum_{j \in \mathcal{N}(\mathbf{x}_k)} u_{ij}$ , where  $\mathcal{N}(\mathbf{x}_k)$  is the set of neighboring pixels, e.g. a square window centered on  $\mathbf{x}_k$ . Note that sFCM reduces to FCM when  $(p, q) = (1, 0)$ .

## III. THE CLUSTER VALIDITY INDEX

For the CV application, we are interested in aggregation functions that map a collection  $\mathbf{u}$  in  $[0, 1]$  to a value in  $[0, 1]$ , formally a vector  $\mathbf{u} = {}^t(u_1, \dots, u_c) \mapsto \mathcal{A}(\mathbf{u})$ . Among the frequently used aggregation operators, one finds the class triangular norms ( $t$ -norms) and conorms ( $t$ -conorms). They have been introduced to characterize the general multi-valued logic *AND* and *OR* operations and are widely used in fuzzy logic and fuzzy set theory to implement conjunctive and disjunctive operators respectively. A  $t$ -norm is a commutative, associative and monotone function  $\top$  having for neutral element 1. Alternatively, its dual operator,

the  $t$ -conorm, is a commutative, associative and monotone function  $\perp$  having for neutral element 0. Examples of norms couples are given in Table I for two operands, see [6] for a large survey, including parameterized families. The dual couple is generally associated with a strong fuzzy negation defined as  $N(v) = 1 - v$  and mentioned as the triple  $(\top, \perp, N)$ .

Table I  
USUAL TRIANGULAR NORM COUPLES

<b>Standard</b>	$u_1 \top_S u_2 = \min(u_1, u_2)$
	$u_1 \perp_S u_2 = \max(u_1, u_2)$
<b>Algebraic</b>	$u_1 \top_A u_2 = u_1 u_2$
	$u_1 \perp_A u_2 = u_1 + u_2 - u_1 u_2$
<b>Łukasiewicz</b>	$u_1 \top_L u_2 = \max(u_1 + u_2 - 1, 0)$
	$u_1 \perp_L u_2 = \min(u_1 + u_2, 1)$

Given a fuzzy clustering  $(U, V)$  of  $X$ , each  $u_{ik}$  defines the similarity of the object  $\mathbf{x}_k$  to the prototype  $\mathbf{v}_i$ , or the degree to which  $\mathbf{x}_k$  satisfies the  $i^{\text{th}}$  group description. Using the membership vector  $\mathbf{u}_k = {}^t(u_{1k}, \dots, u_{ck})$  ( $k^{\text{th}}$  column of  $U$ ), the standard  $t$ -conorm (max operator) is commonly used to select the most appropriate group the object should be assigned to. However, such an exclusive partitioning is not efficient because lower values interact with the greatest value, meaning that  $\mathbf{x}_k$  satisfies more than one group description. Therefore, it is adequate to aggregate the values in  $\mathbf{u}_k$  in order to assess to which extent  $\mathbf{x}_k$  belongs or not to several clusters and how many. We propose to use the fuzzy OR operator (fOR- $l$  for short) defined in [7] for supervised classification with reject options purpose. This operator evaluates degrees of similarity at a given order ( $l$ ) by combination of  $t$ -norms. Let  $\mathcal{P}$  be the power set of  $C = \{1, 2, \dots, c\}$  and  $\mathcal{P}_l = \{A \in \mathcal{P} : |A| = l\}$  where  $|A|$  denotes the cardinality of subset  $A$ , then the fOR- $l$  associates to  $\mathbf{u}_k$  a single value  $\perp^l(\mathbf{u}_k) \in [0, 1]$  defined by:

$$\perp^l(\mathbf{u}_k) = \bigwedge_{i=1, \dots, c}^l u_{ik} = \bigtop_{A \in \mathcal{P}_{l-1}} \left( \bigwedge_{j \in C \setminus A} u_{jk} \right) \quad (3)$$

It must be viewed as some kind of generalization of the “ $l^{\text{th}}$  highest” value,  $l \in C$ . Using standard  $t$ -norms,  $\perp^l(\mathbf{u}_k)$  is exactly the  $l^{\text{th}}$  highest element of  $\mathbf{u}_k$  (see [7]).

Validating the provided clustering  $(U, V)$  of  $X$  consists in assessing whether the resulting partition reflects the data structure or not. Due to the unsupervision, the user does not have any prior knowledge on  $c$  whereas it is required by clustering algorithms such as FCM. Most of works on CV focus on the number of clusters problem. A reliable CVI for FCM must consider both *compactness* and *separation* of the resulting fuzzy  $c$ -partition  $U$ . If only the former is considered, the best partition consists of as many clusters as points. On the other hand, if only the latter is considered,

the best partition consists of a single cluster. So it is well established that a CVI has to combine both measures to be efficient. Separation measures generally take distances between centroids into account but it is seldom sufficient to interpret the geometrical structure of the data, and therefore the separation between clusters, see [3] for examples on the limitations of these approaches.

We propose to use, for each point  $\mathbf{x}_k$ , two measures that overcome these drawbacks: a fuzzy *overlap* measure which evaluates the degree of overlap of a specified number ( $l$ ) of fuzzy clusters and a fuzzy *separation* measure which indicates the degree of overlap of the most probable cluster, i.e. the one corresponding to the highest membership degree, with respect to the  $c - 1$  other ones. A low value of this latter measure will denote a large separation of the most probable cluster from the others. Despite the ability to deal with overlapping clusters is now considered to be a major quality of a CVI [8], most of works are based on an intuitive representation of *overlap*. An overlap measure between  $l$  fuzzy clusters for each point  $\mathbf{x}_k$  of  $X$  described by its membership vector  $\mathbf{u}_k$  can be obtained by (3). By successively computing  $\perp^l(\mathbf{u}_k)$  for different values of  $l$ , we get a combination of  $l$ -order overlap degrees for  $\mathbf{x}_k$ . In order to determine the overall degree of overlap for a given point, it is adequate to determine which orders induce high overlap. The most satisfied order(s) can be quantified by the fuzzy disjunction of the  $l$ -order overlap measures ( $l = 2, c$ ), so we define the overall overlap measure of a point  $\mathbf{x}_k$  by:

$$O_{\perp}(\mathbf{u}_k, c) = \bigwedge_{l=2, c}^1 \left( \bigwedge_{i=1, c}^l u_{ik} \right) \quad (4)$$

In [2], Bezdek and Pal show that inter cluster separation plays a more important role in cluster validity than diameters. We propose to introduce such a measure by quantifying the fuzzy separation of each point  $\mathbf{x}_k$  with  $\perp^l(\mathbf{u}_k)$ , which is the overlap measure within one fuzzy cluster, i.e. its separation from the other fuzzy clusters, since  $\mathbf{u}_k$  sums up to one. This aggregation corresponds to the fuzzy disjunction of membership degrees for a given point  $\mathbf{x}_k$ , which selects the most probable cluster. We define the fuzzy separation of  $\mathbf{x}_k$  with respect to the  $c$  clusters as:

$$S_{\perp}(\mathbf{u}_k, c) = \bigwedge \left( \underbrace{\bigwedge_{i=1, c}^1 u_{ik}, \dots, \bigwedge_{i=1, c}^1 u_{ik}}_{c-1 \text{ times}} \right) \quad (5)$$

Since a small value of the overlapping degree  $O_{\perp}(\mathbf{u}_k, c)$  and a large value of the separation one  $S_{\perp}(\mathbf{u}_k, c)$  indicates that the highest membership degree of  $\mathbf{x}_k$  corresponds to a well separated and not overlapping cluster, we finally define the  $l$ -order *Overlap and Separation Index* (OSI) taking values

in  $[0, 1]$  as the average value of the ratios of both measures:

$$OSI_{\perp}(c) = \frac{1}{n} \sum_{k=1}^n \frac{O_{\perp}(\mathbf{u}_k, c)}{S_{\perp}(\mathbf{u}_k, c)} \quad (6)$$

Running sFCM for  $c$  in a range  $[c_{min}, c_{max}]$  and selecting the value  $c_{best}$  which minimizes (6) allows to segment an image in  $r_{best}$  corresponding homogeneous regions.

#### IV. COLOR IMAGE SEGMENTATION RESULTS

Various studies have shown that many color spaces proposed for computer graphic applications are not well adapted to image processing. As pointed out in [9], a convenient representation should yield distances and provides independence between chromatic and achromatic components. For this reason and comparison purpose, we use the CIE-lab color space, see [10] for details and comparisons. We test the proposed segmentation approach on the Berkeley segmentation database [11] which consists in 300 color images. For each image, a set of benchmark results provided by human observers is given. Performance is evaluated by computing the consistency of the segmentations against the human ones using some classical criteria:

- the *Probabilistic Rand Index* (PRI) which counts the number of pixels whose labels are the same for both segmentations,
- the *Variation of Information* (VoI) which averages the conditional entropy of one segmentation given the other one,
- the *Global Consistency Error* (GCE) which evaluates to which extent one segmentation can be viewed as a refinement of another.

Many unsupervised image segmentation algorithms have

Table II  
PERFORMANCES ON THE BERKELEY DATABASE (BOLD VALUE IS THE BEST SCORE).

Method	PRI	VoI	GCE	rank
1) Mean-shift [12]	0.755	2.477	0.259	5.33
N-cuts [13]	0.722	2.932	0.218	5.33
2) CTM [14]	0.756	2.464	<b>0.176</b>	2.66
FCR [10]	<b>0.761</b>	2.440	0.242	3.66
3) sFCM + XB	0.631	2.412	0.219	5
sFCM + K	0.642	2.325	0.211	3.66
sFCM + OSI <sub>⊥s</sub>	0.751	<b>2.283</b>	0.198	<b>2.33</b>

been proposed so far. Since comparison cannot be done with all of them, we choose to compare the proposed method to three couples of methods:

- 1) Two algorithms that are widely used in the segmentation literature, which complexity is comparable to the CV approach: the *Mean-shift* [12] and *N-cuts* [13]. Mean-shift is a robust effective mode-seeking algorithm based on nonparametric kernel density estimation. The optimal number of modes, depending on some resolution parameters, is automatically performed so segmentation consists

in associating the pixels to the closest local mode. N-cuts (*Normalized-cuts*) is a graph-theoretic method taking each pixel as a node and connecting each pair of pixels by an edge so that the weight on the edge reflects the likelihood that the two pixels belong to the same region. Segmentation is obtained by cutting the edges less than an optimal value, solving an eigenvalue problem. For both algorithms, internal parameters are set to their default values.

- 2) Two more sophisticated and somehow new methods: the *Compression-based Texture Merging* [14] and *Fusion of Clustering Results* [10]. CTM is based on local features assumed to be drawn from a mixture of (possibly degenerate) Gaussians. Component parameters are estimated using the lossy minimum description length and the segmentation is the one that minimizes the number of bits needed to code the segmented data, subject to a given distortion. FCR fuses segmentation maps obtained through the k-means clustering of histograms computed on various color spaces.

- 3) sFCM associated with two popular CVis: the Kwon (K) and Xie-Beni (XB) indexes, see [3] for definitions. For the CV approach, sFCM is used with a square window  $7 \times 7$  for the neighborhood and  $(p, q) = (0, 2)$  so that the smoothing effect is large enough to remove small pixel variations and the computational time is reasonable.

Results reported in Table II show that the proposed method (sFCM + OSI<sub>⊥s</sub>) obtains the overall best mean rank, the best score for VoI, and it outperforms the other CV approaches (group 3) whatever the performance measure. It also gives better results than the non clustering methods (group 1) for two measures. Compared to the more highly sophisticated methods (group 2), it gives quite satisfactory in-between scores. Note that for sFCM associated with the presented CVI, only the results obtained using the standard norms which aggregate cautiously are given. Using algebraic and Łukasiewicz norms sometimes produce over- and under-segmented results respectively, because they involve input values interactions and are Archimedean ( $\top(u_1, u_1) < u_1$ ) whereas standard ones are not.

#### V. CONCLUSION

A non-parametric segmentation method based on a cluster validity index, provided fuzzy partitions of pixels in a given color space is proposed. It is associated to a spatial version of the fuzzy c-means algorithms in order to take into account neighboring pixels. The index allows to select a number of clusters that are as separated and non overlapping as possible and to use it to segment an image in homogeneous regions. Results show the efficiency of the proposed approach, compared to other unsupervised segmentation algorithms. We intend to adapt it to the color space selection problem to compare it to other sophisticated methods. In a future work, we will intensively study how mathematical properties of triangular norms affect the selected partition, and therefore the segmentation.



Figure 1. Images segmented by sFCM+OSI<sub>LS</sub>.

## REFERENCES

- [1] J. C. Bezdek, *Pattern Recognition with fuzzy objective function algorithm*. Plenum Press, 1981.
- [2] J. Bezdek and N. Pal, "Some new indexes of cluster validity," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 23, no. 3, pp. 301–315, 1998.
- [3] D. Kim, K. Lee, and D. Lee, "On cluster validity index for estimation of the optimal number of fuzzy clusters," *Pattern Recognition*, vol. 37, no. 10, pp. 2009–2025, 2004.
- [4] N. Pal and J. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 3, pp. 370–379, 1995.
- [5] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized Medical Imaging and Graphics*, vol. 30, no. 1, pp. 9–15, 2006.
- [6] E. Klement and R. Mesiar, *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Elsevier, 2005.
- [7] L. Mascarilla, M. Berthier, and C. Frélicot, "A k-order fuzzy or operator for pattern classification with k-order ambiguity rejection," *Fuzzy Sets and Systems*, vol. 159, no. 15, pp. 2011–2029, 2008.
- [8] M. Bouguessa, S. Wang, and H. Sun, "An objective approach to cluster validation," *Pattern Recognition Letters*, vol. 27, no. 13, pp. 1419–1430, 2006.
- [9] J. Angulo and J. Serra, "Color segmentation by ordered mergings," in *IEEE International Conference on Image Processing*, 2003.
- [10] M. Mignotte, "Segmentation by fusion of histogram-based k-means clusters in different color spaces," *IEEE Transactions on Image Processing*, vol. 17, no. 5, pp. 780–787, 2008.
- [11] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int. Conf. Computer Vision*, vol. 2, 2001, pp. 416–423.
- [12] D. Comaniciu and P. Meer, "Mean-shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.
- [13] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.
- [14] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy coding and compression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, pp. 1546–1562, 2007.