Spatial-region classification by Min-Connected algorithm for unsupervised segmentation

Rachid Alaoui[1]*, Abdeslam Jakimi[2] and Lahcen Elbermi[3]

[1]*Laboratory LISQ,Faculty of Sciences Dhar-Mahraz, P.O. Box 1796, Atlas-Fes, Morocco*
[2]*Operational Research & Computer Team, My Ismail University, FSTE, B.P 509, Boutalamin, Errachidia, Morocco*
[3]*Operational Research & Computer Team, My Ismail University, FSTE, B.P 509, Boutalamin, Errachidia, Morocco*
*Corresponding author E-mail: ra_alaoui@yahoo.fr

Abstract

This work lies within the scope of color image segmentation by spatial-region classification. The spatial distribution of objects in each region of image is difficult to be identified when the region are non-connected clusters. A standard of color identification by conventional methods of segmentation remains weak for capturing the spatial dispersion of the various objects of the same color region. We propose to apply a spatial classification to characterize geographical connected sets that represent the same regions. The originality of this paper lies in our new min-connected algorithm which is derived from a spatial-color compactness model. Our work is a hybrid segmentation that takes into account the distribution of colors in the color space and the spatial location of colors in the image plane. Experimental tests on synthetic and real images show that our technique leads to promising results for segmentation.

Keywords: *Color image, Segmentation, Clustering, K-means, Spatial connected.*

1   Introduction

Image segmentation is a principal issue Pattern Recognition involves, which is also a fundamental computer vision technology [1]. Image segmentation is a kind of technology and process of dividing an image into homogenous regions, through which we separate the image into several areas with their own features and obtain (pick up) the targets we are interested in, and these areas, not intersected with each other. The earlier research work in the segmentation field was mainly concentrated on gray images. Color image is the best approach to the world, and then are more advantages in color image processing than in gray image. With the development of computer, color image segmentation is being paid more and more attention. To segment a color image, we should select an appropriate color space firstly, and then segmentation method fit to this color space is to be used.

Here, we approach the segmentation problem by using clustering techniques. The HSV color space is fundamentally different from the widely known RGB color space since it separates out intensity (luminance) from the color information (chromaticity). We assume that image pixels are points in HSV color space in such a way that points of similar color, belonging to an image region, will appear in this space as dense cluster. Therefore, the image regions will be presented by dense clusters in the color space or feature space. After, these clusters are then mapped back to the original spatial domain to produce a segmentation of the image.

When image segmentation is viewed as a clustering process, many approaches to image segmentation have widely been used to solve the segmentation by connected color region (e.g., K-means [2]-[3], fuzzy c-means [4], Gaussian mixture model [5]-[6]-[7]). When the problem of color image segmentation is approached by a color classification, it is however a gap in the sense of spatial representation of different objects for each color region, a classification based on color alone is not sufficient. In this paper, to solve this problem by using the concept of spatial connected, we propose a new algorithm (noted Min-Connected) which initialed by clustering technique. We use the K-means model for initialization process, because it is a fast algorithm [3], and effectively has been used in a wide variety of important practical situations. Our new algorithm Min-Connected localized automatically different objects for each non connected region. The originality of this paper lies in the hybrid clustering which is derived from a spatial-color model. Our algorithm takes into account both the distribution of colors in the color space and the spatial location of colors in the image plane.

This paper is organized as follows. Section 2 describes the segmentation by a colorimetric connected model, and section 3 describes the segmentation by a spatial connected model. The implementation and experiments of our new algorithm is shown in section 4. Finally, section 5 is devoted to concluding remarks.

2   Unsupervised segmentation by a colorimetric connected model

Clustering based on the K-Means algorithm is a widely used region segmentation method which, however tends to produce color connected regions.

2.1   Pretreatment of images

In general color images are affected by a noise that increases with intensity. This noise implied a bad segmentation. However, it is necessary to integrate a filter to efficiently alleviate over-segmentation, and the results of segmentation are close to human perception. We use the median filter to eliminate noise and unnecessary details of labeled images. Due to these advantages, this filter is not only useful for processing large amount of image data but also suitable for embedded systems which have few computational resources.

2.2   K-Means

K-Means clustering attempts to subdivide samples consisting of feature values into a set of clusters based on the distances between the samples. Features that are close to each other will be grouped together. The method is quite fast, simple and has been applied and shown to be useful in many applications. To briefly explain the concept of the k-means algorithm, we suppose the observations are $x_1,....,x_L$. The goal of the K-Means algorithm is to partition the observations into K groups with mean such that

$$D(K) = \sum_{i=1}^{L} \underset{1 \leq j \leq K}{Min} (x_i - \overline{x_j})^2 \quad (1) \quad \text{is minimized.}$$

K-Means clustering works by iterating the following two steps until convergence: (1) assign each observation to the closest cluster-mean, and (2) update the cluster mean to the centroid of all observations assigned to it in the previous step.

K-Means algorithm is a quick sense of execution time. In addition, this algorithm can be used easily in the process of segmentation. However, k-means ignore the distribution of objects in each region.

2.3   Determination of optimal clusters number based on cluster validity index

As we know, if the number of clusters K is correctly selected, good clustering result can be achieved; otherwise, data points cannot be grouped into appropriate clusters, and image segmentation cannot be performed appropriately. To this end, we use the cluster validity index.

Cluster validity index V was proposed in 1999 by Siddheswar Ray [8]. This index enhanced performances when compared with the previous studies that used other validity indexes such as the partition coefficient [9-10], all of which have a common goal to find the clustering which results in compact clusters which are well separated. The Davies-Bouldin index [9], for example, is a function of the ratio of the sum of within-cluster scatter to between-cluster separation. The objective is to minimize this measure as we want to minimize the within-cluster scatter and maximize the between-cluster separation. Bezdek and Pal [11] have given a generalization of Dunn's index [12].

The optimal cluster number $K*$ is determined by using two measures:

The first measure is intera-cluster distance (intera) which is simply the distance between a point and its cluster centre and we take the average of all of these distances, defined as :

$$\text{intra} = \frac{1}{N} \sum_{j=1}^{K} \sum_{x \in \chi_I} \|v_i - x_j\|^2$$

The second measure is inter-cluster (inter) distance, we calculate this as the distance between cluster centres, and take the minimum of this value, defined as

$$\text{inter} = Min_{i \neq j} \left\| v_i - v_j \right\|^2$$

Since we want both of these measures to help us determine if we have a good clustering, we must combine them in some way. The obvious way is to take the ratio, defined as:

$$\text{validity} = \frac{\text{intra}}{\text{inter}}$$

Since we want to minimize the intra-cluster distance and this measure is in the numerator, we consequently want to minimize the validity measure. We also want to maximize the inter-cluster distance measure, and since this is in the denominator, we again want to minimize the validity measure. Therefore, the clustering which gives a minimum value for the validity measure will tell us what the ideal value of K is in the k-means procedure.

3   Unsupervised segmentation by a spatial connected model

After segmenting of the color image I by K-means algorithm, we can be extract K color regions noted $(R_k)_{1 \leq k \leq K}$

3.1   The notion of spatial connected

The concept of spatial connected is related to the existence of paths in a graph: for a pixel belongs to region R_k, there exists a path of pixels belongs R_k to reach any other pixel of R_k ?

Each region is not necessarily connected, this implies that K-Means is weak to capture the spatial dispersion of objects of the same color region. A spatial classification will be cited in the following to detect objects in each color region.

The Figure 1 gives a simple example of the connectedness. We see that the region A (green color) is connected but the region B (blue color) is non-connected.

Fig1: Two regions A and B

We defined an object as a connected component of pixels of the same color region. Our goal is locate objects directly for each region. So we assume that the proposed algorithm starts from an initial state where the image is segmented by K-Means. Each pixel of the image is linked to its neighbors to determine the rate of homogeneity. We then establish a list of related objects associated with each color region by assigning each pixel of the image a label that will be stored in a table of the same size as the image.

3.2   Segmentation by the Min-Connected algorithm

Extracting and labeling of various disjoint and connected components in an image is central to many automated image analysis applications. For characterizing the objects in each region, we scans the image and groups its pixels into components based on pixel connectivity, i.e. all pixels in a connected component share similar pixel intensity values and are in some way connected with each other. The connected component labeling (object) is to assign the same label to pixels belonging to the same connected component. So we construct a table of equivalence (denoted T) of the same size as the image that manages the equivalence of labels that appear in the sequential scan. The algorithm scans the

image bottom to top and from left to right and seeks to distinguish the objects for each region R_k. We labeled Each pixel P(i, j) in image I as number of labeling T (i, j) (denoted Etiq(P (i, j))), the algorithm gathers the pixels of the same color to maintain the connectivity of objects, by allocating the same label in T. To label the pixels, we must compare it with neighboring pixels. For this step, we chose to use the 4-neighborhood (in Fig. 2 we assigned the red color to 4-neighborhood of current pixel P(i,j).

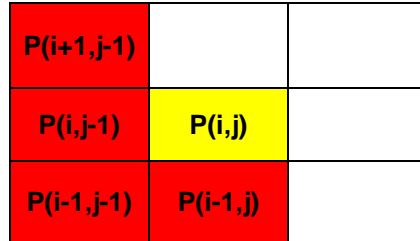| P(i+1,j-1)		
P(i,j-1)	P(i,j)	
P(i-1,j-1)	P(i-1,j)	

Fig2: A 4-neighborhood of current pixel P(i,j)

To locate objects in each region, we start with initial labeling, we gives 0 for the pixels of region and -1 outside. When scanning the image pixel by pixel, we give a new label for a pixel that belongs to a new object detected. In general the geometric shape of an object is non-convex. This problem implies that the neighborhood of pixels running along the scan may be more than one number labeling strictly positive, or an object must have the same number labeling. So we must calculate the minimum strictly positive number of pixels labeling belongs to 4-neighborhood of current pixel in the sequential scan, mathematically, to build our Min-connected algorithm, we should formally explicit the minimum number strictly positive of two numbers x and y.

A simple form is: $\varphi(x, y) = A(x, y) \times Max(x, y) + B(x, y) \times Min(x, y)$ for $(x, y) \in (IR^- \times IR^{*+}) \cup (IR^{*+} \times IR^-) \cup (IR^{*+} \times IR^{*+})$

with A and B are : $A(x,y) = \begin{cases} 0 & if \ (x, y) \in IR^{*+} \times IR^{*+} \\ 1 & if \ (x, y) \notin IR^{*+} \times IR^{*+} \end{cases}$ and $B(x, y) = \begin{cases} 1 & if \ (x, y) \in IR^{*+} \times IR^{*+} \\ 0 & if \ (x, y) \notin IR^{*+} \times IR^{*+} \end{cases}$

A pseudo code of our new algorithm is as follows:

Min-Connected Algorithm

Input :A region R_k of the image I (segmentation by k-means algorithm).

Output : locate every object in the region R_k

```
 eti=1 ;
 1. Initialization of table of labelling:
 For every pixel P(i,j) from Rk ,T(i,j)=0 ;
 For every pixel P(i,j)  from  I − Rk , T(i,j)=-1 ;
 2. Scanning the image pixel by pixel
  If the current pixel P(i,j) belongs to Rk  :
 {
   If a new object is detected
   {
     T(i,j)=eti ; eti++ ;
   }
   else
   {
     m = φ(φ(T(i + 1, j − 1), T(i, j − 1)), φ(T(i − 1, j − 1), T(i − 1, j)));
     If ∃q a pixel neighbour of P(i,j) with etiq(q)>m
     {
      Change all the numbers labelling etiq(q) from T to  m;
       T(i,j)=m;
     }
     else  { T(i,j))=m;}
     }
 }
```

4   Main results

In this paper, our program is complied by Microsoft visual C++ 6.0. Our new algorithm is tested on a variety of images. Fig 3 and Fig 4 show the segmentation results by objects (Min-Connected Algorithm) of two original images "shoes" and "cats". The connected components extracted by our new algorithm in the two images are effectively the objects in each region segmented by K-Means. From the experiment, we see that our new algorithm get satisfactory extraction of objects from regions images.
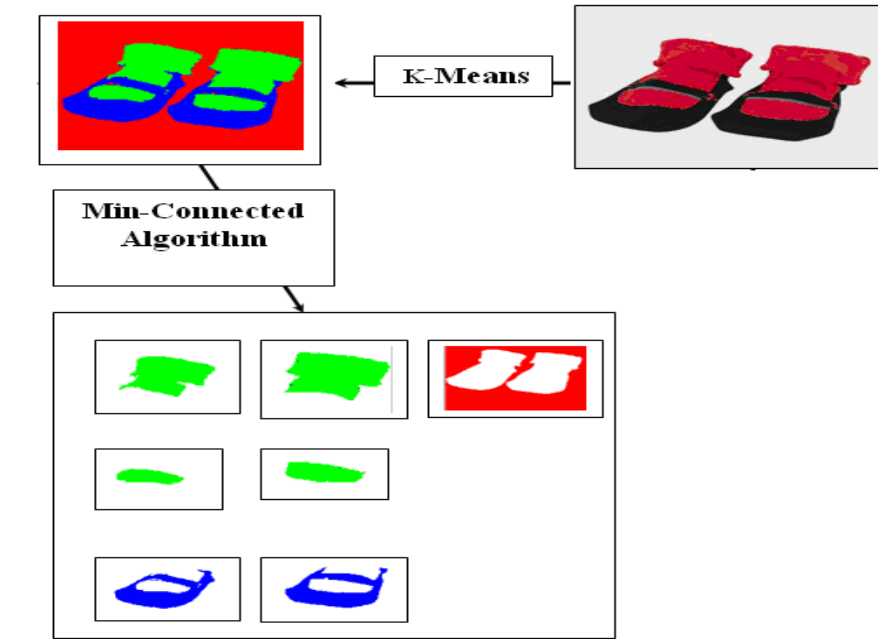

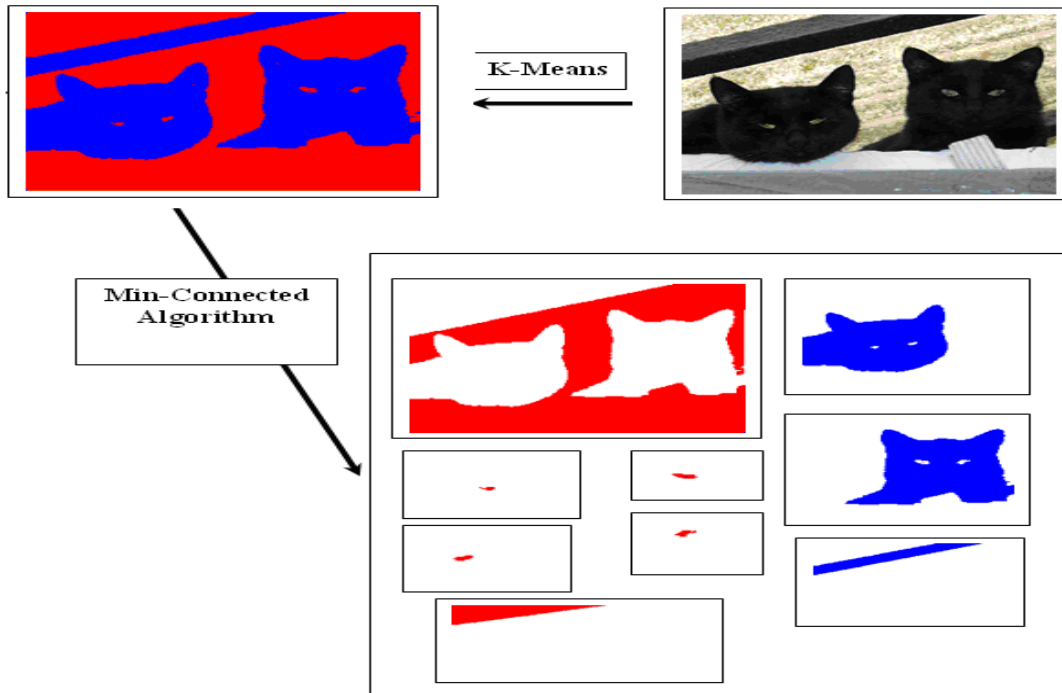Fig 3: Segmentation by objects using Min-Connected Algorithm

Fig 4: Segmentation using by objects using Min-Connected Algorithm

5   Conclusion

In this paper, color image segmentation has been considered as a pixel spatial classification problem. Our Min-connected algorithm consists in constructing the classes of pixels by searching objects into region. The construction of the classes is achieved by the spatial clustering approach, which is well-designed to identify non-convex objects in the region. We have shown the effectiveness of our algorithm by segmenting a synthetic image which contains objects with non-convex and strong overlapping geometric objects. Moreover, we have tested our algorithm with natural images in order to check that the simultaneous analysis of the spatial and color properties of the objects is a relevant strategy. This work gives room to several possibilities for further improvement. First, we can develop a retrieval system of color images based into objects extraction. Second, the estimation of semantic of image is not always well adapted for the image segmentation by region classification. It would be interesting to study the relationships between the objects.

References

[1]   ZHANG Yu-jin. Image project (media),image analysis. Beijing:Tsinghua University Press,2005.

[2]   A.K.Jain, M.N.Murty, and P.J.Flynn. Data clustering: a review. ACM Computing Surveys, 31(3):264-323, 1999.

[3]   Tse-Wei Chen, Yi-Ling Chen, and Shao-Yi Chien, "Fast image segmentation based on K-Means clustering with histograms in HSV color space," Proceedings of IEEE International Workshop on Multimedia Signal Processing (MMSP2008), Cairns, Queensland, Australia, Oct 2008, pp. 322–325.

[4]   Clark MC, Hall LO, Goldgof DB, Clarke LP, Velthuizen RP, Silbiger MS. MRI segmentation using fuzzy clustering techniques. IEEE Eng Med BioI (1994) 13:730-42.

[5]   REDNER,R.A. and WALKER,H.F., 1984. Mixture densities, maximum likelihood and the EM algorithm, SIAM Rev., 26,195–239.

[6]   A. Lachkar, R. Benslimane, M. Daoudi "Extraction de Motifs de Zellige Arabo-Moresque par une nouvelle Méthode de Segmentation Non-Supervisée d'Images Couleurs"coresa2004, 2004

[7]   A.Lachkar, R.Benslimane, L.D'Orazio and E.Martuscelli "A system for textile design patterns retrieval. Part I: Design patterns extraction by adaptive and efficient color image segmentation method"Journal of the Textile Institute, Vol.97, July , pp.301-312,2006

[8]   S. Ray and R.H. Turi, Determination of Number of Clusters in K-means Clustering and Application in Color Image Segmentation, In Proc. of the 4th Intern. Conf. on Advances in Pattern Recognition and Digital Techniques, India, 1999..

[9]   D.L. Davies and D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Machine Intell. Vol. 1, pp. 224-227, 1979.

[10]   N.R. Pal and J.C. Bezdek, On cluster validity for the fuzzy c-means model, IEEE Trans. Fuzzy Systems, vol. 3, pp. 370-379,1995.

[11]   J.C. Bezdek and N.R. Pal, Some new indexes of cluster validity, IEEE Trans. Syst.Man. Cybern., vol. 28, pp. 301-315, 1998.

[12]   J.C. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, J. Cybern.,vol. 3,pp. 32-57,1973.