Neurocomputing 120 (2013) 235–249

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Color segmentation by fuzzy co-clustering of chrominance color features

Madasu Hanmandlu [a,1], Om Prakash Verma [b], Seba Susan [a,*], V.K. Madasu [c]

[a] IIT Delhi, New Delhi, India
[b] Delhi Technological Univ., Delhi, India
[c] Queenland Univ., Brisbane, Australia

ARTICLE INFO

Article history:
Received 19 December 2011
Received in revised form
19 September 2012
Accepted 22 September 2012
Available online 2 April 2013

Keywords:
Fuzzy Co-clustering
Object membership
Feature membership
Validity measure
Bacterial Foraging
Color segmentation

ABSTRACT

This paper presents a novel color segmentation technique using fuzzy co-clustering approach in which both the objects and the features are assigned membership functions. An objective function which includes a multi-dimensional distance function as the dissimilarity measure and entropy as the regularization term is formulated in the proposed fuzzy co-clustering for images (FCCI) algorithm. The chrominance color cues a^* and b^* of CIELAB color space are used as the feature variables for co-clustering. The experiments are conducted on 100 natural images obtained from the Berkeley segmentation database. It is observed from the experimental results that the proposed FCCI yields well formed, valid and high quality clusters, as verified from Liu's *F*-measure and Normalized Probabilistic RAND index. The proposed color segmentation method is also compared with other segmentation methods namely Mean-Shift, NCUT, GMM, FCM and is found to outperform all the methods. The bacterial foraging global optimization algorithm gives image specific values to the parameters involved in the algorithm.

1. Introduction

The segmentation of color images is a potential area of research due to its practical significance in various fields. Image segmentation partitions the image into regions/segments such that pixels belonging to a region are more similar to each other than those belonging to different regions. Clustering is a well known approach for segmenting images. It strives to assess the relationships among patterns of the data set by organizing them into groups or clusters such that patterns within a cluster are more similar to each other than those belonging to different clusters. Many algorithms for both hard and fuzzy clustering have been developed to achieve this purpose. In hard clustering, data is divided into crisp clusters, where each data belongs to exactly one cluster. In fuzzy clustering, the data points can belong to more than one cluster, and associated with each of the points are membership grades that indicate the degree to which the data points belong to the different clusters. Clustering in the color domain gives improved segmentation results since color components carry more information than the gray scale components.

Several techniques have been proposed in the field of color segmentation. Histogram based segmentation of color [1] is one of the existing techniques. But it does not guarantee contiguity of the resulting regions. Edge detection based techniques [2] pose the difficulty of determining the boundary of an image due to the ambiguity of the response of a weak edge. Recently, Arbelaez et al. in [3] have proposed a hierarchical segmentation obtained from the output of a contour detector which overcomes the difficulties of weakly linked boundaries. In [4] color segmentation by region growing and merging is investigated. One drawback of the conventional region growing technique is the selection of the seed point and the order in which regions grow or merge. In [5], the problem of seed selection is solved by using the relaxation labeling technique which yields satisfactory results. Recent techniques for region growing use automated seed selection process as in [6] which uses a fuzzy similarity and fuzzy distance based approach. In [7], after the region growing of similar color, Markov Random Fields (MRF) are applied to improve the results. However, it is observed that some homogeneous regions may get disconnected due to the MRF process. Blobworld [8], a popular image segmentation and retrieval algorithm groups pixels into regions by modeling the joint distribution of color texture and position features by a mixture of Gaussians with parameters being decided by the expectation maximization algorithm. However, the resulting blobs may not contain all the details of objects and also may not distinguish an object which is not visually distinct. Further an iterative post-processing step is required to correct the

* Corresponding author. Tel.: +91 98 68850209.
E-mail addresses: mhmandlu@iitd.ac.in (M. Hanmandlu), opverma.dce@gmail.com (O.P. Verma), sebasusan@indiatimes.com, seba_406@yahoo.in (S. Susan), v.madasu@uq.edu.au (V.K. Madasu).
[1] Senior Member IEEE.

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.09.043

mis-alignment of object boundaries. Mean-Shift filtering [9] and Graph partitioning [10,11] methods and their hybrids [12] perform clustering in feature–space and are found to be effective for color segmentation. But they are very sensitive to the parameters like color bandwidth (Mean-Shift) and the threshold edge length (Graph method). Neural network based approaches [13,14] for image segmentation like Competitive Learning Neural Network (CLNN) and Self Organizing of Kohonen Feature Map (SOFM) avoid complex programming but usually consume a lot of training time. Other significant works on image segmentation include: Watershed technique[15] based on the morphological watershed transform, segmentation using the K-nearest neighbor (K-NN) technique [16] which is sensitive to the choice of reference sample and JSEG [17]—a segmentation algorithm based on color and texture. Various combinations of popular segmentation algorithms like region merging and graph partitioning [18], mean-shift and region merging [19], watershed and Kohonen SOM [20] have been suggested together with their advantages. In [21], color segmentation is carried out by applying a set of fuzzy if-then-rules on 200 fixed color samples. The Fuzzy C-Means (FCM) clustering method, a popular choice for color segmentation has been investigated in the works of [22]. The results are quite good but for the computational complexity and sensitivity to the initialization. Several variants of FCM are summarized in [14]. In [23] fuzzy set theory and maximum fuzzy entropy principle are used to convert the image to the fuzzy domain and a Space Scale filter is used to analyze the homogeneity histogram to find the appropriate segments. Fuzzy co-clustering algorithm with its dual fuzzy (object and feature) membership functions was originally derived for document clustering, examples being FCCM, FCoDoK [24,25] and robust versions PFCC [26], RFCC[27]. The co-clustering done so far on images [28,29] is limited to indexing of images for Content based image retrieval (CBIR) in which low level semantic features derived from image histogram are the feature variables for clustering.

In this paper the Fuzzy co-clustering approach is adapted for the segmentation of natural images. An algorithm for the Fuzzy Co-clustering of images (FCCI) is developed by incorporating the distance between each feature data point and the feature cluster center as the dissimilarity measure and the entropies of the objects and features as the regularization terms in the objective function. To prove the effectiveness of our approach we apply the FCCI algorithm for the segmentation of color images with successful results. Some preliminary work on color segmentation of histo-pathological images is reported in [30] and this serves as a precursor to the main work. The CIELAB color space is favored for our experiments due to the wide range of colors possible and its closeness to the human perception system [31], and its chrominance color vector $\{a^*,b^*\}$ is proved to be the best feature combination for the segmentation task [32]. It is found from the experimental results that the color segmentation results obtained by the proposed technique are of high quality with respect to both Liu's evaluation measure and NPR index which are global segmentation evaluation measures and also outperforms over other popular color segmentation methods. The choice of the number of clusters for the experiment is determined in a novel manner by plotting Xie and Beni's cluster validity [33] as the number of clusters is increased and by checking for the first local minima of the curve. The resulting segmentation offers a good tradeoff between color difference and human perception.

The organization of the paper is as follows: The Fuzzy Co-Clustering algorithm for Images (FCCI) is introduced in Section 2. While minimizing the objective function in FCCI, Bacterial Foraging is adapted for global learning of parameters. Later in Section 3, the proposed algorithm is applied for the color segmentation together with the state of the art comparisons. Finally conclusions from over-all results are given in Section 4.

2. Fuzzy co-clustering algorithm for images

2.1. Motivation for the algorithm and related work

Co-clustering simultaneously clusters both objects and features together [24]. This provides two membership functions: the partition or object membership function and the ranking or the feature membership function. The latter serves to filter out the relevant features only during the computation of the object membership function and thus solves the problem of sparseness of data by reducing the dimensionality. The co-clustering algorithm is thus suited to applications with large dimensions and is found to be apt for our experiments on multi-feature color images. The problem of outliers is also minimized by using feature membership function [26]. The problem with using only the feature memberships is that it may lead to coincident/overlapping clusters therefore highlighting the need for both feature and object memberships. Further we include the distance function of feature data points from the feature cluster centroids in the co-clustering process to create richer co-clusters than other fuzzy co-clustering algorithms. The inclusion of the distance factor in the degree of aggregation reduces the optimization problem to a minimization one. In this work co-clustering is integrated with the Fuzzy approach with a view to obtain distinct clusters [24,26]. Both the object and feature memberships in the proposed method are fuzzy, i.e. the object membership is calculated when different clusters compete for a data point and feature memberships are defined when different features compete for a cluster. Thus we have two constraints on the two fuzzy memberships (object and feature memberships) in our method.

Therefore the aim is to have a co-clustering algorithm with the following advantages:

1. It must be insensitive to initialization and form distinct clusters. (Fuzzy clustering)
2. It should perform well in high dimensions and provide well defined clusters. (Co-clustering)
3. It should minimize the impact of outliers to improve the accuracy of co-clustering. (Ranking/Feature memberships)
4. Its objective function should integrate the distance measure of input features w.r.t. feature centroids into the entropy regularization framework.
5. It must be reasonably fast enough.

Several maximum entropy clustering algorithms and their variants are available in the literature [34,35]. One such approach of interest to the present work is the variant of FCM which props on entropy regularization [36]. It involves the minimization of the following objective function:

$$J_{FCM} = \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} Dist(x_i, p_c) + T_U \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} \log u_{ci} \tag{1}$$

subject to the constraint

$$\sum_{c=1}^{C} u_{ci} = 1,\ u_{ci} \in [0,1], \forall i = 1,\ldots,N \tag{2}$$

where the symbols C,N represent the number of clusters and data points respectively, u_{ci} the fuzzy membership function, T_U the weight factor in the entropy term, $Dist(x_i,p_c)$ the dissimilarity term equal to the square of the Euclidean distance between pixel x_i and cluster center p_c.

The first term in the R.H.S of (1) denotes the effective squared distance; the second term is the entropy which serves as a regulating factor during the minimization process. The proposed approach aims at co-clustering in the entropy framework of FCM. For this we begin by replacing the distance function $Dist(x_i,p_c)$with