퍼지컬러 모델을 이용한 컬러 데이터 클러스터링 알고리즘[1]

Color Data Clustering Algorithm using Fuzzy Color Model

김대원 이광형
한국과학기술원 전산학과

Dae-Won Kim Kwang H. Lee
Department of Computer Science, AITrc, KAIST
E-mail : dwkim@if.kaist.ac.kr

ABSTRACT

The research interest of this paper is focused on the efficient clustering task for an arbitrary color data. In order to tackle this problem, we have tried to model the inherent uncertainty and vagueness of color data using fuzzy color model. By taking a fuzzy approach to color modeling, we could make a soft decision for the vague regions between neighboring colors. The proposed fuzzy color model defined a three dimensional fuzzy color ball and color membership computation method with the two inter-color distance measures. With the fuzzy color model, we developed a new fuzzy clustering algorithm for an efficient partition of color data. Each fuzzy cluster set has a cluster prototype which is represented by fuzzy color centroid.

Key words : Fuzzy color model, Fuzzy clustering, Cluster centroid

I. Introduction

Color is the one of the most important features in our lives. It is not easy task to effectively describe color in our language even though color can be considered as a simple and intuitive object[1][2]. The research interest of this paper is focused on the color clustering problem. For a given set of color data and the number of clusters, the objective is to partition the color set into homogenous color sub-partition. This kind of color clustering task can be widely used for a variety of applications, for example, color image segmentation. The difficulties of this research include not only the lack of correct color model which can describe the uncertain characteristics of color, but also the lack of efficient clustering algorithm which can deal with the vague color data.

In order to effectively partition the color data set, fuzzy cluster analysis technique would be the best choice due to the ability of dealing with color uncertainty. But most of previous fuzzy clustering algorithms have been designed for crisp pattern data, thus we need to develop a new algorithm that can tackle the fuzzy data clustering. We applied the fuzzy color model to devise a new approach in fuzzy clustering algorithm[3]. Color is a qualitative feature and we have to make a soft decision in color clustering activity. Thus each color cluster set is prototyped by color centroid value that is represented by fuzzy color ball. By minimizing the defined evaluation criteria, we could obtain the sub-optimal convergence in an iterative color clustering.

II. Research Background

2.1 Cluster Analysis for Color Data

Data clustering, often called cluster analysis, organizes a collection of data patterns which is usually represented as a vector of measurements or a point in a multidimensional space into clusters based on a kind of similarity measure[4]. The clustering task classifies a given collection of unlabeled data into homogeneous groups without any labeled patterns. Most of clustering researches to deal with color data have handled the color as a general crisp element without the consideration of uncertainty and vagueness. Thus popular clustering algorithms with a conventional color, e.g. RGB color, represented in crisp color space have been just used.

The practical techniques for cluster analysis are K-Means clustering or FCM clustering for RGB color representation. With these approaches, however, we cannot handle the color uncertainty problem. Yang and Liu recently proposed a class of fuzzy c-numbers clustering procedures for fuzzy data where high dimensional fuzzy vector data are handled[5]. But the fuzzy data with conical fuzzy vector is somewhat theoretical and is not suitable for color representation. Thus we developed an effective fuzzy clustering algorithm for color data with the help of the fuzzy

[1] This work was supported by the Korea Science and Engineering Foundation through the Advanced Information Technology Research Center(AITrc).

color model that can explain the color uncertainty and relative color difference formula.

2.2 The Concept of Fuzzy Color Model

One of the major factors affecting color description is that color depends on neighboring color stimuli in the observer's visual field This color description can be understood by computing perceived color difference among given colors. The observer may judge the color difference relatively and subjectively. This is the reason none of the many color difference formula that have been proposed in the literature over the past several decades is considered as a sufficiently adequate solution of the problem. Most of color models including RGB and CIELAB have aimed at the crisp color representation and color difference formula. The simple use of the color representation does not explain the similar perception and visual confusion of certain colors.

In order to tackle these problems we proposed a new color model based on fuzzy set theory[3]. Color can be described as a membership degree to a specific color, and the notion of color membership resolves the vague boundaries between colors and relative color acceptability. The fuzzy color, denoted by $\tilde{c}_i \in \tilde{C}$,is described with three dimensional ball-type shape on CIELAB color space. The color ball has two attributes: (1) center value that is the center point of fuzzy color ball on CIELAB color coordinate (2) JND means a just noticeable distance which is the distance from the center to a boundary of a given color. Figure 1 depicts the fuzzy color ball and its membership function shape. For a given color element x, the membership $\mu_{\tilde{c}_i}(x)$ of x to fuzzy color $\tilde{c}_i$ is obtained with the distance computation. The fuzzy color has its own center value $center_i$ and JND value jnd_i that build three-dimensional ball-shaped representation. If a color x is within JND distance it strongly belongs to that color and has membership degree 1.0. If color x is out of the range of JND, the membership degree is relatively determined by comparing with neighbor colors. The left and right shape of fuzzy membership function is determined based on the distance result between fuzzy colors.

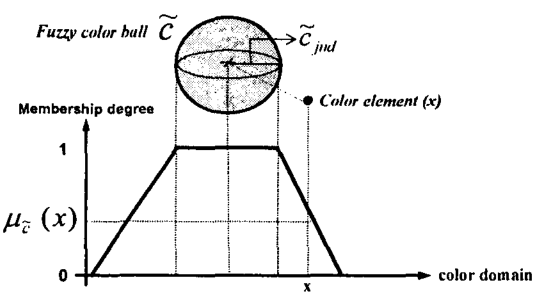

Fig 1. Fuzzy color ball and its membership shape

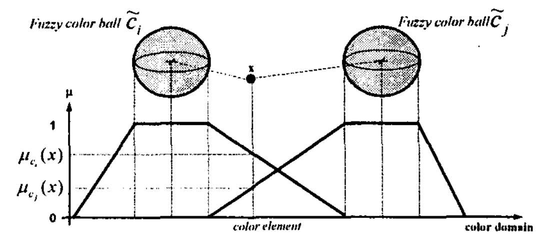

Fig 2. Membership computation among fuzzy colors

With the distance computation measures, we defined a formal fuzzy color model. In this paper, we omit the detailed definition of fuzzy color ball. The below equation shows the color membership computation method to a specific fuzzy color.

$$\mu_{\tilde{c}_i}(x) = \begin{cases} 1.0 & \text{if } \delta(\tilde{c}_i, x) \le jnd_i \\ 0.0 & \text{if } \delta(\tilde{c}_j, x) \le jnd_j, (i \ne j) \\ \left(\sum_{j=1}^{|\tilde{C}|} \frac{\delta(\tilde{c}_i, x)}{\delta(\tilde{c}_j, x)} \right)^{-1} & \text{otherwise} \end{cases} \quad (1)$$

where δ-function means the distance between a fuzzy color and a color element. Figure 2 shows the membership computation between two fuzzy colors. In this case, color element x doesn't belong to any fuzzy colors and is located to in the middle of fuzzy colors. The color membership degree to two fuzzy colors are calculated with the above equation 1.

III. Proposed Fuzzy Clustering Method

3.1 Formulation of Color Clustering Problem

The goal of our research is to cluster the given color data into homogeneous color groups. To accomplish this, fuzzy clustering technique is selected as a major algorithm approach because the major feature 'color' of the data pattern has uncertain and vague properties. Thus fuzzy approach has an advantage over the hard clustering algorithms to handle these problems.

The following terms and notations are used to describe the problem formulation. The given color element data are $x_1, \ldots, x_n \in X$ where each x_j is a color element represented as three dimensional *Lab* value on CIELAB color space. The algorithm objective is to cluster a collection of given color elements into c homogeneous groups represented as fuzzy sets (F_i, $i=1,\ldots,c$) with similar color characteristics. The fuzzy cluster set can be written as equation 2.

$$F_i = \{(x_1, \mu_{F_i}(x_1)), (x_2, \mu_{F_i}(x_2)), \ldots, (x_n, \mu_{F_i}(x_n))\} = \sum_{j=1}^{n} \mu_{F_i}(x_j)/x_j \quad (2)$$

The collection of fuzzy cluster set is denoted by $F=\{F_1, F_2, \ldots, F_c\}$. Each fuzzy cluster is defined by its centroid, denoted by c_i, which is represented by the proposed fuzzy color model. The pattern matrix M which is handled by clustering algorithm is represented as a $c \times n$ pattern matrix. An element p_{ij} in matrix M means that a color element x_j is classified to a fuzzy cluster F_i.

The equation 3 shows the evaluation function of the proposed approach. The objective is to minimize the evaluation function $J(F)$ for a given pattern matrix.

$$J(F) = \sum_{i=1}^{c} \sum_{j=1}^{n} \left[\mu_{F_i}(x_j)\right]^2 \delta(c_i, x_j) \quad (3)$$

where $\delta(c_i,x_j)$ is the distance between the color element x_j and the centroid c_i of the fuzzy cluster set F_i. The goal is to iteratively improve a sequence of sets of fuzzy clusters $F(1)$, $F(2)$,.. until $F(t)$ is found such that no further improvement in $J(F)$ is possible. In general, the design of membership function and centroid prototype are the most important problems.

3.2 Initialization of Cluster Analysis

Before we describe the detailed clustering algorithm, we discuss how to obtain the initial partition of color cluster set. In cluster analysis, initialization plays an important role. According to the initial starting point, the clustering algorithm might terminate at different clustering results. There is no general agreement about a good initialization scheme. Two most popular initialization techniques are (1) using the first c pattern data (2) using c elements randomly from pattern set. In this paper, we propose a novel initial selection method based on the notion of fuzzy color. It is simple and intuitive.

The basic idea is as follows. For a color element $x_j \in X$, we compute the matching degree to pre-defined thirteen reference fuzzy colors which is denoted by c_R. Reference fuzzy colors are selected from both the major color family in Munsell color wheel (5R, 5YR, 5Y, ..., 5P, and 5RP) and the L-axis components (white, black, gray) of CIELAB color space. Figure 3 shows the overall initial selection process. Color elements in pattern space is input value to the color decision system, and the output is a list of fuzzy colors where the fuzzy colors are sorted by the maximum matching score. The first c fuzzy colors are chosen as the initial fuzzy cluster centroids.

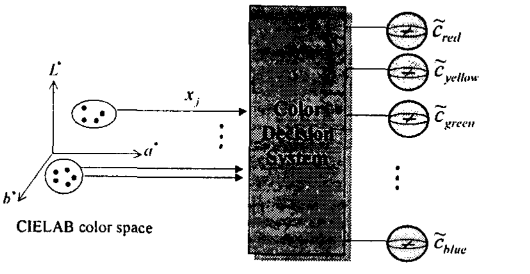

Fig 3. Initial selection of fuzzy cluster centroid

For a given color element x_j, the matching score is computed by considering the membership degree to all reference fuzzy colors. Each reference fuzzy color $c_i \in C_R$ has two additional attributes denoted by $winner(i)^{count}$ and $winner(i)^{element}$. The $winner(i)^{count}$ means how many color elements belong to the given reference fuzzy color c_i.

$$winner_i^{counter} \leftarrow winner_i^{counter} + 1 \quad \text{if } \mu_{c_i}(x_j) \geq \mu_{c_k}(x_j) \ (\forall c_k \in C_R) \quad (4)$$

After the classification decision process, we build a sorted list of thirteen fuzzy colors according to their $winner(i)^{count}$ in a descending order. From the sorted list we select the first c (number of clusters) fuzzy colors which have a larger matching counter than other colors. We don't need to focus on the whole thirteen colors in the list, only c fuzzy colors are enough for the candidates for cluster centroid.

3.3 Fuzzy Clustering using Fuzzy Color

The previous fuzzy clustering algorithm including FCM was designed only to deal with crisp data element. This means they can be regarded as a fuzzy clustering algorithm for crisp data. They did not consider any mechanisms about how to handle fuzzy data, especially color characteristics, even though the fuzzy clustering methodology is very appropriate to efficiently deal with the vague boundaries between fuzzy color data. Cluster analysis can be explained in the viewpoint of pattern matrix M. By updating the pattern matrix iteratively in each sequence step, clustering algorithm goes on the convergence condition. The typical matrix M has a $c \times n$ matrix form, where each cluster centroids c_i represent the row attribute of M, and each data element x_j stands for the column of matrix. In this situation all color data and cluster centroids have crisp representation, hence the conventional approach has a limitation to account for the uncertainty and vagueness in color clustering.

We tried to represent the fuzziness and vagueness resolution method into pattern matrix. Proposed approach is to replace the typical point-prototype crisp centroid with a new cluster centroid represented by the fuzzy color. Figure 4 depicts the skeleton of the proposed pattern matrix. The column of matrix contains each color elements x_j, and the row elements represent the centroid c_i of each fuzzy cluster F_i which is shown as three dimensional fuzzy color ball. Fuzzy color centroid has its own center element and JND value that are a specific properties of a given fuzzy cluster. This makes it possible to cope with the clustering decision problem in the vague region between neighbor fuzzy color clusters. With the initialization step to obtain the initial fuzzy color centroids c_i, the membership computation of each element $\mu(x_j)$ in pattern matrix is carried out. The

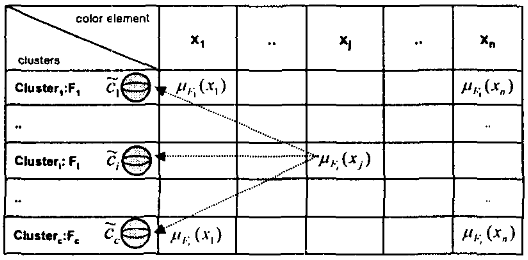

Fig 4. Membership update in pattern matrix

membership degree of color element x_j to the fuzzy cluster F_i is computed by the relative distance between neighbor fuzzy color centroids. After this cluster initialization process, fuzzy color centroids and the membership degrees of matrix elements are iteratively updated until no improvements are found.

3.4 Skeleton of Proposed Algorithm

As mentioned in earlier section, the objective is to obtain the optimal partition $F=\{F_1, F_2,.., F_c\}$ for a given color elements $x_1,...,x_n \in X$ and number of clusters c by minimizing the evaluation function $J(F)$. The algorithmic step can be described as followings:

Step 1. With a given pre-determined number of clusters (c) and the color data elements ($x_1,...,x_n \in X$), run the proposed cluster initialization method to obtain the candidates of fuzzy color centroids.

Step 2. Create the initial c fuzzy color centroids, $c_1(t),...,c_c(t)$ where t means the iteration step (initially $t=0$) with the candidates obtained in Step 1. The newly created fuzzy color centroid c_i is defined as follows:

$$c_i = \langle center_i, jnd_i \rangle$$

$$center_i \leftarrow winner_i^{element}$$

$$jnd_i \leftarrow jnd_k \;\; s.t. \; \{k \mid \forall c_k \in C_R, \wedge_k \delta(c_k, center_i)\}$$

where δ-function means the distance between a color element and a fuzzy color, and C_R represent the universe of discourse of the predefined thirteen reference fuzzy colors.

Step 3. Update the membership degree of fuzzy cluster sets F_i(t+1) by the following procedure. For each color element x_j, compute as the followng.

$$\mu_{F_i}(x_j)(t+1) = \begin{cases} 1.0 & if \; \delta(c_i, x_j) \le jnd_{c_i} \\ 0.0 & if \; \delta(c_k, x_j) \le jnd_{c_k}, (i \ne k) \\ \left(\sum_{k=1}^{c} \frac{\delta(c_i, x_j)}{\delta(c_k, x_j)} \right)^{-1} & otherwise \end{cases}$$

where the column sum constraint should be satisfied.

Step 4. Update the fuzzy color centroid c_i of each fuzzy cluster. The jnd_i of each fuzzy color centroid c_i is updated in a similar way of step 2. The $center_i$ value is computed by the following procedure.

$$center_i(t+1) = \frac{\sum_{j=1}^{n} \mu_{F_i}(x_j)(t) \cdot x_j}{\sum_{j=1}^{n} \mu_{F_i}(x_j)(t)}$$

Step 5. If $|F_i(t+1) - F_i(t)| < \varepsilon$ for all $F_i \in F$, where ε is a small positive constant, then halt because it's believed that algorithm has reached at convergence; otherwise, t ← t+1 and go to step 3.

IV. Conclusion

In this paper we discussed the color clustering problem. To successfully partition the color pattern data, we adopted the fuzzy color model that can describe the vagueness of natural colors. With the definition of three dimensional fuzzy color ball and inter-color distance measures, we could calculate the color membership degree for a given arbitrary color element. In order to effectively deal with color data in clustering, we adopted a fuzzy cluster analysis. Because the fuzzy clustering makes a soft decision in each iteration through the use of membership functions, it may be the best technique in processing the fuzzy color data. We developed a new fuzzy clustering algorithm with fuzzy color model. The key idea was to exploit the fuzzy color centroids. Each fuzzy color centroid can help to calculate the membership degree of each color data.

As further works, we would like to study the automatic determination of the number of clusters and centroid initialization technique. sThe hardening issue is also very important because it is closely related to the defuzzification of fuzzy cluster sets.

V. References

[1] J.T. Luke. The Munsell Color System: A language for color. Fairchild Publications, 1996.

[2] P. Zelanski and M.P. FisherM.P. Color. Calmann and Kng Ltd, 1998.

[3] 김대원, 이광형, "퍼지컬러 모델을 이용한 컬러의 소속정도를 결정하는 방법에 관한 연구", 정보과학회 2002 춘계학술대회 발표논문집(B), pp. 262-264, 2002.

[4] A.K. Jain, et al. "Data Clustering: A Review", *ACM Computing Surveys*, Vol. 31, No. 3. pp. 264-323, 1999.

[5] M.S. Yang and H.H. Liu. "Fuzzy Clustering Procedures Conical Fuzzy Vector Dsata", *Fuzzy sets and systems*, Vol. 106, pp. 189-200, 1999.

[6] R.S. Berns. Principles of Color Technology. Wiley-Interscience Publication, New York, 2000