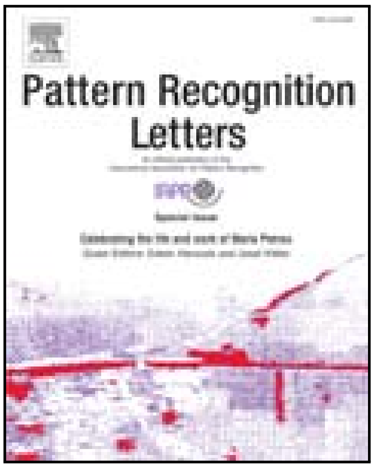
Accepted Manuscript



Diverse Fuzzy c-Means for Image Clustering

Lingling Zhang, Minnan Luo, Jun Liu, Zhihui Li, Qinghua Zheng

PII: S0167-8655(18)30290-3
DOI: 10.1016/j.patrec.2018.07.004
Reference: PATREC 7230

To appear in: *Pattern Recognition Letters*

Received date: 26 April 2018
Revised date: 6 June 2018
Accepted date: 2 July 2018

Please cite this article as: Lingling Zhang, Minnan Luo, Jun Liu, Zhihui Li, Qinghua Zheng, Diverse Fuzzy c-Means for Image Clustering, *Pattern Recognition Letters* (2018), doi: 10.1016/j.patrec.2018.07.004

Pattern Recognition Letters

Authorship Confirmation

Please save a copy of this file, complete and upload as the "Confirmation of Authorship" file.

As corresponding author I, Minnan Luo, hereby confirm on behalf of all authors that:

1. This manuscript, or a large part of it, has not been published, was not, and is not being submitted to any other journal.
2. If presented at or submitted to or published at a conference(s), the conference(s) is (are) identified and substantial justification for re-publication is presented below. A copy of conference paper(s) is(are) uploaded with the manuscript.
3. If the manuscript appears as a preprint anywhere on the web, e.g. arXiv, etc., it is identified below. The preprint should include a statement that the paper is under consideration at Pattern Recognition Letters.
4. All text and graphics, except for those marked with sources, are original works of the authors, and all necessary permissions for publication were secured prior to submission of the manuscript.
5. All authors each made a significant contribution to the research reported and have read and approved the submitted manuscript.

Signature Minnan Luo Date 2018/4/24

List any pre-prints:

Relevant Conference publication(s) (submitted, accepted, or published):

Justification for re-publication:

Graphical Abstract (Optional)

To create your abstract, please type over the instructions in the template box below. Fonts or abstract dimensions should not be changed or altered.

Diverse Fuzzy *c*-Means for Image Clustering

Lingling Zhang, Minnan Luo, Jun Liu, Zhihui Li, Qinghua Zheng

Image clustering is a key technique for better accomplishing image annotation and searching in large image repositories. Fuzzy c-means and its variations have achieved excellent performance on image clustering because they allow each image to belong to more than one cluster. However, these methods neglect the relations between different image clusters, and hence often suffer from the "cluster one-sidedness" problem that redundant centers are learned to characterize the same or similar image clusters. To this issue, we propose a diverse fuzzy c-means for image clustering via introducing a novel diversity regularization into the traditional fuzzy c-means objective. This diversity regularization guarantees the learned image cluster centers to be different from each other and to fill the image data space as much as possible. An efficient alternative optimization algorithm is exploited to address the proposed diverse fuzzy c-means objective, which is proved to converge to local optimal solutions. Experiments on synthetic and six image datasets demonstrate the effectiveness of the proposed method as well as the necessity of the diversity regularization.

ACCEPTED MANUSCRIPT

Research Highlights (Required)

To create your highlights, please type the highlights against each `\item` command.

It should be short collection of bullet points that convey the core findings of the article. It should include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point.)

- A diverse fuzzy c-means is introduced for better accomplishing image clustering.
- Diversity term makes the image cluster centers to cover as many clusters as possible.
- A smooth relaxation is used to replace the non-smooth and non-convex diversity term.

Pattern Recognition Letters
journal homepage: www.elsevier.com

Diverse Fuzzy c-Means for Image Clustering

Lingling Zhang[a,b], Minnan Luo[a,b,**], Jun Liu[a,c], Zhihui Li[d], Qinghua Zheng[a,c]

[a]School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China
[b]Ministry of Education Key Lab For Intelligent Networks and Network Security, Xi'an, China
[c]National Engineering Lab for Big Data Analytics, Xi'an Jiaotong University, Xi'an, China
[d]School of Computer Science and Engineering, University of New South Wales, Sydney, Australia

ABSTRACT

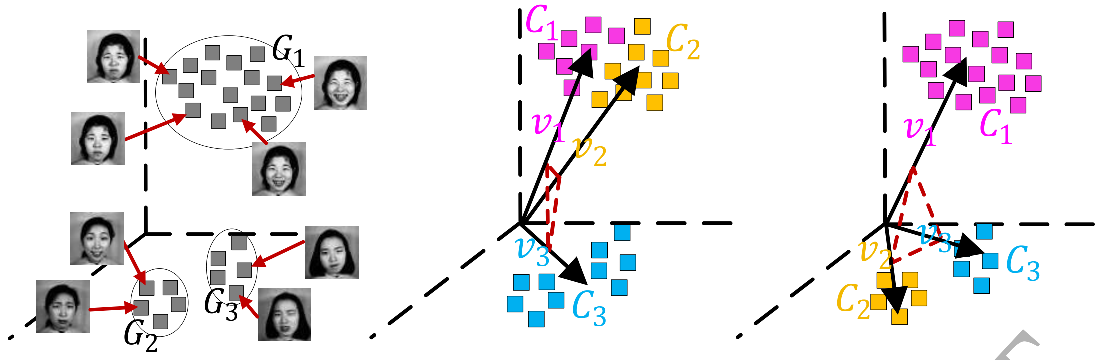
Image clustering is a key technique for better accomplishing image annotation and searching in large image repositories. Fuzzy c-means and its variations have achieved excellent performance on image clustering because they allow each image to belong to more than one cluster. However, these methods neglect the relations between different image clusters, and hence often suffer from the "cluster one-sidedness" problem that redundant centers are learned to characterize the same or similar image clusters. To this issue, we propose a diverse fuzzy c-means for image clustering via introducing a novel diversity regularization into the traditional fuzzy c-means objective. This diversity regularization guarantees the learned image cluster centers to be different from each other and to fill the image data space as much as possible. An efficient optimization algorithm is exploited to address the diverse fuzzy c-means objective, which is proved to converge to local optimal solutions and has a satisfactory time complexity. Experiments on synthetic and six image datasets demonstrate the effectiveness of the proposed method as well as the necessity of the diversity regularization.

1. Introduction

Image clustering aims to partition a set of images into clusters such that the images in the same cluster share some similarity whereas images from different clusters are as dissimilar as possible (Yang et al., 2010; Li et al., 2017; Yuan et al., 2017). With the explosion of image data, image clustering becomes an effective data analysis tool that helps to better accomplish image annotation, image searching, and other applications over the large image repositories (Chang et al., 2017a; Lai et al., 2014). For example, Gordon *et al.* (Gordon et al., 2003) utilized the image clustering preprocessing to improve the speed and performance of contented-based image retrieval; Jia *et al.* (Jia et al., 2008) adopted image clustering to obtain the meaningful conceptual labels for annotating personal photo albums more effectively; Kennedy *et al.* (Kennedy and Naaman, 2008) clustered the landmark images to generate some typical images for representing diverse views of the landmarks.

For image clustering, each image always includes few visual objects with rich semantic information (Liu et al., 2018a,b), which not only belongs to a single image cluster in most cases. Therefore, soft c-partition, where c denotes the prefixed number of clusters, is well suited for this task because it assigns each image to more than one cluster simultaneously. Among various techniques of soft c-partition, fuzzy c-means derived from fuzzy logic is one of the most significant ones for its simplicity and effectiveness (Dunn, 1973; Bezdek, 2013). In the past decades, fuzzy c-means and its variations (Zhu et al., 2009; Hathaway and Hu, 2009; Yang and Nataliani, 2017) have achieved good results on image clustering and other areas. For example, Wang *et al.* (Wang et al., 2014) utilized the approximation optimization to solve a stochastic gradient based fuzzy clustering (SGFC) algorithm for large image clustering; Bhat (Bhat, 2014) introduced possibility fuzzy c-means clustering (PFCM) to reduce the influences from the expression, pose, makeup, illumination for face recognition; Ding *et al.* (Ding et al., 2010) presented a novel fuzzy c-means algorithm with cluster-center-free reformulation, which achieved good perfor-

**Corresponding author: Tel.: +86-15829263196;
e-mail: zhanglingling@stu.xjtu.edu.cn (Lingling Zhang), minnluo@xjtu.edu.cn (Minnan Luo), liujun@xjtu.edu.cn (Jun Liu), zhihuilics@gmail.com (Zhihui Li), qhzheng@xjtu.edu.cn (Qinghua Zheng)



(a) Original Image Data (b) Fuzzy c-means (c) Diverse fuzzy c-means

Fig. 1: Left: the original dataset includes 26 face images from three groups G_1, G_2 and G_3. Center: the conventional fuzzy c-means leads to "cluster one-sidedness" problem because two similar cluster centers are learned to represent one group G_1. Right: the diverse fuzzy c-means address "cluster one-sidedness" problem by controlling the learned centers to be different from each other.

mance on handwritten image clustering. However, previous studies on fuzzy c-means learn the image cluster centers just by minimizing the sum of squared distances between each image to cluster centers. They ignore the characteristics of image cluster centers as well as the relations between different clusters. Therefore, these methods are easy to cause the "cluster one-sidedness" problem in which few redundant centers correspond to the same or similar image clusters. As shown in Figure 1a, the original face image database consists of 26 gray-scale images posed by 3 females. Three groups denoted by G_1, G_2 and G_3 respectively include 16, 5, and 5 image samples for clustering. Figure 1b demonstrates the partitional result using conventional fuzzy c-means, where the pink, yellow and blue square samples respectively belong to the learned three clusters C_1, C_2 and C_3. Note that three black vectors labeled as $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ denote three learned image cluster centers. Apparently, fuzzy c-means partitions the samples from groups G_1 into clusters C_1 and C_2 according to different expressions, and combines the samples from group G_2 and G_3 as cluster C_3. This algorithm fails to identify groups G_2 and G_3 and learns two similar centers $\mathbf{v}_1$ and $\mathbf{v}_2$ for the larger group G_1. Indeed, the image cluster centers are learned to represent the corresponding image clusters. If the image cluster centers are limited to differ from each other, they will cover the image data space as much as possible, rather than be controlled by the part of data. As a result, controlling the diversity between image cluster centers could effectively alleviate the proposed "cluster one-sidedness" problem during image clustering.

Motivated by the diversity analysis in (Xie et al., 2015; Xie, 2015), we adopt the similar strategy to enhance the diversity of the image cluster centers. Because the angle is invariant to translation, rotation and scaling with positive factors of the two vectors, we leverage the angle between each pair of image centers to measure the pairwise dissimilarity. In this case, we define the diversity metric following the rationale that image cluster centers possess a larger diversity if the pairwise angles between them have a larger mean and a smaller variance (Xie et al., 2015). Specifically, a larger mean implies that these image centers share larger angles on the whole, hence are more different from each other. A smaller variance indicates that these image centers uniformly spread out to different directions and each centroid is evenly different from all other centers. As a result, encouraging the diversity metric to be larger can render the cluster centers to be more diverse and avoid redundant centers effectively. After that, we propose a diverse fuzzy c-means algorithm for image clustering via integrating the mentioned diversity regularization into the traditional fuzzy c-means objective. As shown in Figure 1c, the diverse fuzzy c-means could accurately partition the original image data by expanding the angle mean and shrinking the angle variance over cluster centers $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$. Three contributions of this paper are summarized as follows:

- To encourage the learned image centers to cover as many clusters as possible, we introduce a novel diversity regularization into the conventional fuzzy c-means objective for better image clustering.
- Due to the non-smooth and non-convex of the original diversity regularization, a smooth relaxation is presented to address the proposed optimization problem with alternative iteration algorithm.
- We conduct extensive experiments on synthetic and six image datasets to illustrate the effectiveness and superiority of our method. The experimental results demonstrate that our method consistently outperforms other competitors.

The remainder of this paper is organized as follows. In Section II, we propose diverse fuzzy c-means for image clustering to enhance the discrimination of image cluster centers. Section III focuses on the optimization algorithm to address the proposed diverse fuzzy c-means. Section IV presents some synthetic and real application experiments to illustrate the effectiveness of our method. Conclusions are given in Section V.

2. Diverse Fuzzy c-Means

In the framework of image clustering, the given image dataset is recorded as $X = \{\mathbf{x}_k \in \mathbb{R}^n : k = 1, 2, \cdots, N\}$, where each image $\mathbf{x}_i$ is represented by a n-dimensional feature vector and N refers to the number of images. We aim to partition these images into c fuzzy clusters, where $\mathbf{v}_i \in \mathbb{R}^n$ $(i = 1, 2, \cdots, c)$ denotes the center for the i-th image cluster. The matrix $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_c)$ is a n-by-c matrix consisting of c image cluster centers. In this case, the diverse fuzzy c-means for image clustering is proposed by solving the following objective function:

$$\min_{U,V} \quad J_{m,\lambda}(X;U,V) = \sum_{k=1}^{N} \sum_{i=1}^{c} \mu_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|_2^2 - \lambda \Omega(V). \qquad (1)$$

$$s.t. \quad U \in \left\{ u_{ik} \in [0,1], \forall i,k; 0 < \sum_{k=1}^{N} u_{ik} < N, \forall i; \sum_{i=1}^{c} \mu_{ik} = 1, \forall k \right\}$$

where $m > 1$ is the weighting exponent (fuzzy index) which determines the amount of fuzziness of the resulting partition. $U = (u_{ik})$ is a c-by-n matrix whose column $\mathbf{u}_k = (u_{1k}, u_{2k}, \cdots, u_{ck})^\top \in \mathbb{R}^c$ represents the membership degree to which image $\mathbf{x}_k$ is labeled with $1, 2, \cdots, c$. Apparently, the first term $\sum_{k=1}^{N} \sum_{i=1}^{c} \mu_{ik}^m \|\mathbf{x}_k - \mathbf{v}_i\|_2^2$ is same as the objective function of traditional fuzzy c-means, which assigns each image to each cluster with a certain membership degree and minimizes the sum of least-squared errors between each image and cluster centers. The second term $\Omega(V)$ denotes a novel diversity regularization on image cluster centers V, which controls the learned image cluster centers to be different from each other and to cover as many image clusters as possible. Hyper-parameter λ is a non-negative regularization parameter which makes a trade-off between the fuzzy c-means term and the diversity regularization. When $\lambda = 0$, the objective function (1) turns to the loss measurement of traditional fuzzy c-means, which can easily lead to the "cluster one-sidedness" problem. While $\lambda > 0$, the diversity regularization $\Omega(V)$ is introduced into traditional fuzzy c-means to address the problem.

2.1. Diversity Regularization $\Omega(V)$

In order to measure the diversity of image cluster centers, *i.e.*, how different each image center is from others, we leverage the angle between each pair of image centers since the angle is invariant to translation, rotation and scaling with positive factors (Xie, 2015). Specifically, the angle between image centers $\mathbf{v}_i$ and $\mathbf{v}_j$ is defined to measure the pairwise dissimilarity of centers as follows

$$\theta(\mathbf{v}_i, \mathbf{v}_j) = \arccos\left(\frac{|\mathbf{v}_i^\top \mathbf{v}_j|}{\|\mathbf{v}_i\| \|\mathbf{v}_j\|} \right).$$

Note that we do not care about the orientation of image centers and prefer the angle to be acute or right. If the angle θ is obtuse, we substitute it with $\pi - \theta$. The diversity of the image cluster centers $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_c)$ is defined as

$$\Omega(V) = \Phi(V) - \Psi(V), \qquad (2)$$

where $\Phi(V)$ is the mean of the non-obtuse angles between all pairs of image centers and $\Psi(V)$ is the variance of these angles. Apparently, the diversity of image centers becomes larger while the pairwise angles between centers have a larger mean $\Phi(V)$ and a smaller variance $\Psi(V)$ (Xie, 2015). A larger mean denotes that these image centers share larger angles as a whole, and a smaller variance represents that these image centers uniformly spread out to different directions. Generally, while all angles between any two image centers are closer to $\pi/2$, the diversity value becomes larger. The diversity value reaches the maximum while the whole image centers are vertical to each other. In summary, encouraging the diversity regularization $\Omega(V)$ to be larger can render the image centers to be more diverse and avoid redundant centers effectively.

2.2. Smooth Diverse Fuzzy C-means

Based on the analysis above, we take the diversity of image cluster centers into consideration and formulate the objective function (1) to learns diverse fuzzy c-means for image clustering task. Regarding to a great challenge that the diversity $\Omega(V)$ is non-smooth and non-convex, we follow the strategy used in (Xie, 2015) to relax the objective function. Specifically, let $V = \widetilde{V}\mathrm{diag}(\mathbf{g})$, where $\mathbf{g} = (g_1, g_2, \cdots, g_c)^\top \in \mathbb{R}^c$ and $g_i = \|\mathbf{v}_i\|_2$. In this sense, the L_2 norm of each column in $\widetilde{V} = (\tilde{\mathbf{v}}_1, \tilde{\mathbf{v}}_2, \cdots, \tilde{\mathbf{v}}_c) \in \mathbb{R}^{n \times c}$ is actually 1. According to the definition of the diversity metric, we have $\Omega(V) = \Omega(\widetilde{V})$ and the optimization problem (1) can be reformulated as

$$\min_{U,\widetilde{V},\mathbf{g}} \quad J_{m,\lambda}(X;U,\widetilde{V},\mathbf{g}) = \sum_{k=1}^{N} \sum_{i=1}^{c} u_{ik}^m \|\mathbf{x}_k - g_i \tilde{\mathbf{v}}_i\|_2^2 - \lambda \Omega(\widetilde{V}) \quad (3)$$

$$s.t. \quad U \in M_{fcn}, g_i \geq 0, \|\tilde{\mathbf{v}}_i\|_2 = 1, \forall i = 1, 2, \cdots, c.$$

In contrast to the problem (1), the reformulated optimization problem (3) aims to constrain the magnitude of cluster centers $\tilde{\mathbf{v}}_i (i = 1, 2, \cdots, c)$ to be 1. However, $\Omega(\widetilde{V})$ is still non-smooth and non-convex. To tackle this issue, we follow the work in (Xie, 2015) to derive an efficient smooth lower bound of $\Omega(\widetilde{V})$. The main results can be summarized in the following theorem.

Theorem 1. *(Xie, 2015) Let $\det(\widetilde{V}^\top \widetilde{V})$ be the determinant of the Gram matrix of $\widetilde{V}$, then $0 < \det(\widetilde{V}^\top \widetilde{V}) \leq 1$. Let*

$$\Gamma(\widetilde{V}) = \arcsin\left(\sqrt{\det(\widetilde{V}^\top \widetilde{V})} \right) - \left(\frac{\pi}{2} - \arcsin\left(\sqrt{\det(\widetilde{V}^\top \widetilde{V})} \right) \right)^2, \qquad (4)$$

then $\Gamma(\widetilde{V})$ is a lower bound of $\Omega(\widetilde{V})$; moreover, $\Gamma(\widetilde{V})$ and $\Omega(\widetilde{V})$ have the same global optimal.

According to the Theorem 1, the objective of diverse fuzzy c-means (3) can be reformulated by replacing $\Omega(\widetilde{V})$ with $\Gamma(\widetilde{V})$, *i.e.*,

$$\min_{U,\widetilde{V},\mathbf{g}} \quad J_{m,\lambda}(X;U,\widetilde{V},\mathbf{g}) = \sum_{k=1}^{N} \sum_{i=1}^{c} u_{ik}^m \|\mathbf{x}_k - g_i \tilde{\mathbf{v}}_i\|_2^2 - \lambda \Gamma(\widetilde{V}) \quad (5)$$

$$s.t. \quad U \in M_{fcn}, g_i \geq 0, \|\tilde{\mathbf{v}}_i\|_2 = 1, \forall i = 1, 2, \cdots, c.$$

Note that the smooth approximation $\Gamma(\widetilde{V})$ has been proved to have the following two properties: (1) The gradient of $\Gamma(\widetilde{V})$ is easy to compute; (2) Maximizing $\Gamma(\widetilde{V})$ with projected gradient ascent (PGA) can increase $\Omega(\widetilde{V})$. With these properties, the optimization problem above can be solved with alternating optimization algorithm in the next Section.

Algorithm 1 Diverse fuzzy c-means for image clustering

Input: Image dataset X, cluster numbers c, fuzzy index m, regularization parameter λ, step-size ρ
Initialization: $U^{(0)}, \widetilde{V}^{(0)}, \mathbf{g}^{(0)}$

while halting criterion is not true **do**

Fuzzy c-partition matrix U update:

$$u_{ik}^{(t+1)} = \left[\sum_{j=1}^{c} \left(\frac{\|\mathbf{x}_k - g_i^{(t)}\mathbf{v}_i^{(t)}\|_2}{\|\mathbf{x}_k - g_j^{(t)}\mathbf{v}_j^{(t)}\|_2} \right)^{\frac{2}{m-1}} \right]^{-1} \quad (i = 1,2,\cdots,c;\ k = 1,2,\cdots,N)$$

L_2 norm of each cluster center g_i update: $g_i^{(t+1)} = \max\left\{0, (\tilde{\mathbf{v}}_i^{(t)})^\top \mathbf{h}^{(t+1)}\right\}$ with $\mathbf{h}^{(t+1)} = \frac{\sum_{k=1}^N (u_{ik}^m)^{(t+1)}\mathbf{x}_k}{\sum_{k=1}^N (u_{ik}^m)^{(t+1)}}$ $(\forall i = 1,2,\cdots,c)$;

Normalized image cluster centers $\widetilde{V}$ update: $\widetilde{V}^{(t+1)} = \mathscr{P}(\widetilde{V}^{(t)} + \rho(\nabla\phi(\widetilde{V}^{(t)}; U^{(t+1)}, \mathbf{g}^{(t+1)}) - \lambda G(\widetilde{V}^{(t)})))$;

end while

Output: $U, V = \widetilde{V}\mathrm{diag}(\mathbf{g})$

3. Optimization Algorithm

We first propose the alternative optimization algorithm among $U, \mathbf{g}$, and $\widetilde{V}$ to solve the objective of diverse fuzzy c-means (5) in Section 3.1. And then the convergence and time complexity of the proposed optimization algorithm is analyzed in Section 3.2 and Section 3.3.

3.1. Optimization Procedure

Update U: Given normalized image cluster centers $\widetilde{V}$ and vector $\mathbf{g}$, this step aims to find the fuzzy c-partition matrix U by optimizing

$$\min_{U} \sum_{k=1}^{N}\sum_{i=1}^{c} \mu_{ik}^m \|\mathbf{x}_k - g_i\tilde{\mathbf{v}}_i\|_2^2 \qquad \text{s.t.} \quad U \in M_{fcn} \tag{6}$$

This problem is similar to the objective function of traditional fuzzy c-means. And the fuzzy c-partition matrix U can be updated using

$$u_{ik} = \left[\sum_{j=1}^{c} \left(\frac{\|\mathbf{x}_k - g_i\tilde{\mathbf{v}}_i\|_2}{\|\mathbf{x}_k - g_i\tilde{\mathbf{v}}_j\|_2} \right)^{\frac{2}{m-1}} \right]^{-1} \tag{7}$$

for $i = 1,2,\cdots,c$ and $k = 1,2,\cdots,N$.

Update g: Due to the independence of components in vector $\mathbf{g} = (g_1, g_2, \cdots, g_c)$, we can update g_i separately by minimizing the following convex optimization problems when U and $\widetilde{V}$ are fixed.

$$\min_{g_i} \ h(g_i) = \sum_{k=1}^{N} u_{ik}^m \|\mathbf{x}_k - g_i\tilde{\mathbf{v}}_i\|_2^2 \qquad \text{s.t.} \quad g_i \geq 0. \tag{8}$$

By setting the gradient $\nabla h(g_i)$ equal to zero, *i.e.* $\nabla h(g_i) = \sum_{k=1}^N 2u_{ik}^m (\mathbf{x}_k - g_i\tilde{\mathbf{v}}_i)^\top (-\tilde{\mathbf{v}}_i) = 0$, we have a closed-form solution

$$g_i = \max\left\{0, \tilde{\mathbf{v}}_i^\top \mathbf{h}\right\}, \tag{9}$$

where $\mathbf{h} = \frac{\sum_{k=1}^N u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^N u_{ik}^m}$ for $i = 1,2,\cdots,c$.

Update $\widetilde{V}$: After we have inferred the fuzzy c-partition matrix U and L_2 norm of each image center g_i $(i = 1,2,\cdots,c)$, then we update $\widetilde{V}$ by minimizing the constraint objective function as follows

$$\min_{\widetilde{V}} \sum_{k=1}^{N} u_{ik}^m \|\mathbf{x}_k - g_i\tilde{\mathbf{v}}_i\|_2^2 - \lambda\Gamma(\widetilde{V}) \qquad \text{s.t.} \quad \|\tilde{\mathbf{v}}_i\|_2 = 1, \forall i = 1,2,\cdots,c. \tag{10}$$

Let $\phi(\widetilde{V}; U, \mathbf{g}) = \sum_{k=1}^N u_{ik}^m \|\mathbf{x}_k - g_i\tilde{\mathbf{v}}_i\|_2^2$, then the gradient of $\phi(\widetilde{V}; U, \mathbf{g})$, denoted by

$$\nabla\phi(\widetilde{V}; U, \mathbf{g}) = \left(\frac{\partial\phi(\widetilde{V}; U, \mathbf{g})}{\partial\tilde{\mathbf{v}}_1}, \frac{\partial\phi(\widetilde{V}; U, \mathbf{g})}{\partial\tilde{\mathbf{v}}_2}, \cdots, \frac{\partial\phi(\widetilde{V}; U, \mathbf{g})}{\partial\tilde{\mathbf{v}}_c} \right)^\top,$$

can be calculated by

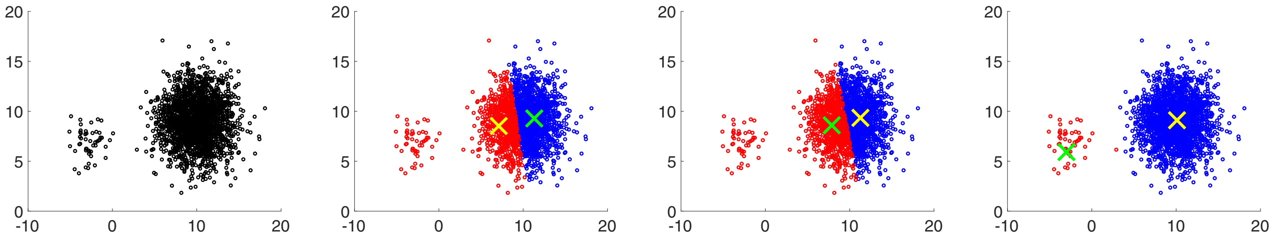
$$\frac{\partial\phi(\widetilde{V}; U, \mathbf{g})}{\partial\tilde{\mathbf{v}}_i} = 2\sum_{k=1}^{N} \mu_{ik}^m g_i (g_i\tilde{\mathbf{v}}_i - \mathbf{x}_k)^\top.$$

Moreover, the gradient of $\Gamma(\widetilde{V})$ with respect to each $\tilde{\mathbf{v}}_i$ can be calculated by

$$\frac{\partial\Gamma(\widetilde{V})}{\partial\tilde{v}_i} = l'(\det(\widetilde{V}^\top\widetilde{V}))\frac{\partial\det(\widetilde{V}^\top\widetilde{V})}{\partial\tilde{v}_i}$$

for $i = 1,2,\cdots,c$, where $l(x) = \arcsin(\sqrt{x}) - (\frac{\pi}{2} - \arcsin(\sqrt{x}))^2$ and the gradient of $l(x)$, $\nabla l(x) = \frac{1}{\sqrt{1-x}}\frac{1}{2\sqrt{x}}[1 + \pi - 2\arcsin(\sqrt{x})]$. As a result, the gradient of $\Gamma(\widetilde{V})$ with respect to V, denoted by $G(\widetilde{V})$, is obtained as

$$\begin{aligned} G(\widetilde{V}) &= \frac{\partial\Gamma(\widetilde{V})}{\partial\widetilde{V}} = \nabla l(\det(\widetilde{V}^\top\widetilde{V})) \times \frac{\partial\det(\widetilde{V}^\top\widetilde{V})}{\partial\widetilde{V}} \\ &= \frac{\sqrt{\det(\widetilde{V}^\top\widetilde{V})}}{\sqrt{1-\det(\widetilde{V}^\top\widetilde{V})}}\left[1 + \pi - 2\arcsin(\sqrt{\det(\widetilde{V}^\top\widetilde{V})})\right] \\ &\quad \times \widetilde{V}(\widetilde{V}^\top\widetilde{V})^{-1}. \end{aligned} \tag{11}$$



(a) 2D Synthetic Data (b) c-means (c) Fuzzy c-means (d) Our Method

Fig. 2: Demonstration on 2D synthetic data. The red and blue circles respectively correspond to two learned clusters, whose cluster centers are represented as the yellow and green crosses.

As a result, the normalized cluster centers $\widetilde{V}$ is updated by projected gradient descent (PGD)(Boyd and Vandenberghe, 2004), *i.e.*,

$$\widetilde{V} = \mathscr{P}(\widetilde{V} + \rho(\nabla\phi(\widetilde{V}; U, \mathbf{g}) - \lambda G(\widetilde{V}))) \tag{12}$$

where $\rho > 0$ is an appropriate step-size and $\mathscr{P}(\cdot)$ denotes the projection to the unit sphere.

In summary, we formulate the alternating optimization algorithm for diverse fuzzy c-means by Algorithms 1, where iterations stop when the minimum bound on the fuzzy c-partition matrix U or the maximum number of iterations is obtained.

3.2. Convergence Analysis

We verify that the proposed iterative procedure in Algorithm 1 converges to the optimal solutions by the following theorem. For better representation, we first recall an important property about the smooth lower bound $\Gamma(\widetilde{V})$ which is utilized to approximate the non-smooth and non-convex term $\Omega(\widetilde{V})$.

Theorem 2. *(Xie, 2015) Let $G^{(t)}$ be the gradient of $\Gamma(\widetilde{V}^{(t)})$ with respect to $\widetilde{V}^{(t)}$ at the t-th iteration. Then there exists $\tau > 0$, such that $\forall \eta \in (0, \tau)$, $\Omega(\widetilde{V}^{(t+1)}) \geq \Omega(\widetilde{V}^{(t)})$ with $\widetilde{V}^{(t+1)} = \mathscr{P}(\widetilde{V}^{(t)} + \eta G^{(t)})$, where $\mathscr{P}(\cdot)$ denotes the projection to the unit sphere.*

With respect to the Algorithm 1 for the proposed diverse fuzzy c-means, the main results of the convergence can be formulated as the following theorem.

Theorem 3. *The alternative updating rules in Algorithm 1 for diverse fuzzy c-means monotonically decrease the objective function (3) in each iteration.*

Proof. For convenience, let us denote

$$\mathscr{L}_m(U, \widetilde{V}, \mathbf{g}) = \sum_{k=1}^{N}\sum_{i=1}^{c} u_{ik}^m \|\mathbf{x}_k - g_i \tilde{\mathbf{v}}_i\|_2^2 - \lambda\Omega(\widetilde{V}) - \gamma\sum_{k=1}^{N}(\sum_{i=1}^{c} u_{ik} - 1) + \eta\sum_{i=1}^{c}(\|\tilde{\mathbf{v}}_i\|_2 - 1). \tag{13}$$

where $\gamma, \eta > 0$ are the Lagrange parameters. With $\widetilde{V}^{(t)}$ and $\mathbf{g}^{(t)}$ fixed, we follow the analysis on the updating of partition matrix U and have

$$U^{(t+1)} = \arg\min_{U \geq 0} L_m(U, \widetilde{V}^{(t)}, \mathbf{g}^{(t)}) \tag{14}$$

with $\sum_{i=1}^{c} u_{ik}^{(t+1)} = 1$ for $k = 1, 2, \cdots, N$. As a result, we obtain

$$\mathscr{L}_m(U^{(t+1)}, \widetilde{V}^{(t)}, \mathbf{g}^{(t)}) \leq \mathscr{L}_m(U^{(t)}, \widetilde{V}^{(t)}, \mathbf{g}^{(t)}). \tag{15}$$

With $U^{(t+1)}$ and $\widetilde{V}^{(t)}$ fixed, by the process of updating $\mathbf{g}$, see Function (9), it is evident to verify that

$$\mathscr{L}_m(U^{(t+1)}, \widetilde{V}^{(t)}, \mathbf{g}^{(t+1)}) \leq \mathscr{L}_m(U^{(t+1)}, \widetilde{V}^{(t)}, \mathbf{g}^{(t)}). \tag{16}$$

In addition, according to Theorem 2, the smooth lower bound of $\Gamma(\widetilde{V})$ can achieve the descend of non-smooth and non-convex diversity regularization $\Omega(\widetilde{V})$, such that $\Omega(\mathscr{P}(\widetilde{V}^{(t)} + \eta G^{(t)})) \geq \Omega(\widetilde{V}^{(t)})$, where $G^{(t)}$ is the gradient of $\Gamma(\widetilde{V}^{(t)})$ with respect to $\widetilde{V}^{(t)}$ at the t-th iteration. Combing the updating of $\widetilde{V}$ in Function (12), we arrive at

$$\mathscr{L}_m(U^{(t+1)}, \widetilde{V}^{(t+1)}, \mathbf{g}^{(t+1)}) \leq \mathscr{L}_m(U^{(t+1)}, \widetilde{V}^{(t)}, \mathbf{g}^{(t+1)}). \tag{17}$$

Based on the Function (15), (16) and (17), we obtain

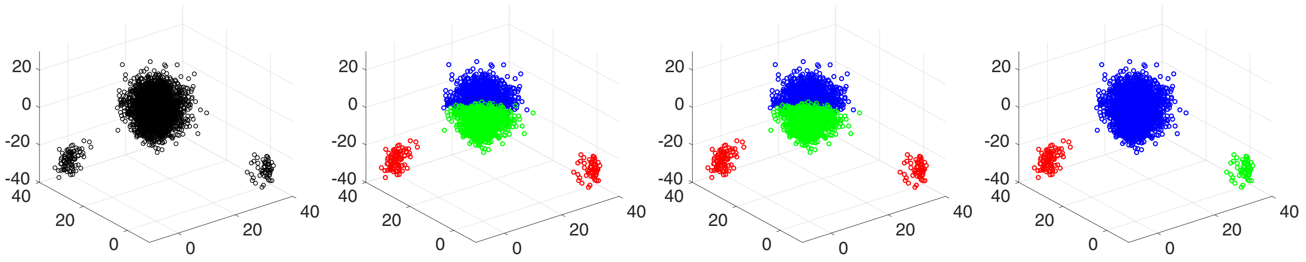
$$\begin{aligned} \mathscr{L}_m(U^{(t+1)}, \widetilde{V}^{(t+1)}, \mathbf{g}^{(t+1)}) &\leq \mathscr{L}_m(U^{(t+1)}, \widetilde{V}^{(t)}, \mathbf{g}^{(t+1)}) \\ &\leq \mathscr{L}_m(U^{(t+1)}, \widetilde{V}^{(t)}, \mathbf{g}^{(t)}) \leq \mathscr{L}_m(U^{(t)}, \widetilde{V}^{(t)}, \mathbf{g}^{(t)}). \end{aligned} \tag{18}$$

Consequently, the objective function of diverse fuzzy c-means monotonically decreases using the updating rules in Algorithm 1 and the proof is completed. □

Based on Theorem 3, we can conclude that the iterative approach in Algorithm 1 converges to local optimal solutions efficiently. In the experiment part as follows, we observe that the proposed alternating algorithm usually converges within no more than 30 iterations, and thus the convergence speed is fast.

3.3. Time Complexity Analysis

We briefly discuss the computational complexity of the proposed Algorithm 1, where three variable U, $\widetilde{V}$ and $\mathbf{g}$ are updated iteratively at each iteration. Specifically, the optimization procedure of U requires $O(Nnc)$, in which N, n and c respectively denote the number of images, the dimension of image feature, and the number of image clusters. When update $\widetilde{V}$, the computing of $\nabla\phi(\widetilde{V}; U, \mathbf{g})$ and $G(\widetilde{V})$ separately needs $O(Nc)$ and $O(nc^2 + c^3)$. After that, it requires $O(Nnc)$ for updating the variable $\mathbf{g}$ with the fixed U and $\widetilde{V}$. Therefore, the total time complexity of the proposed Algorithm 1 is



(a) 3D Synthetic Data (b) c-means (c) Fuzzy c-means (d) Our Method

Fig. 3: Demonstration on 3D synthetic data. The red, blue and green circles respectively correspond to three learned clusters.

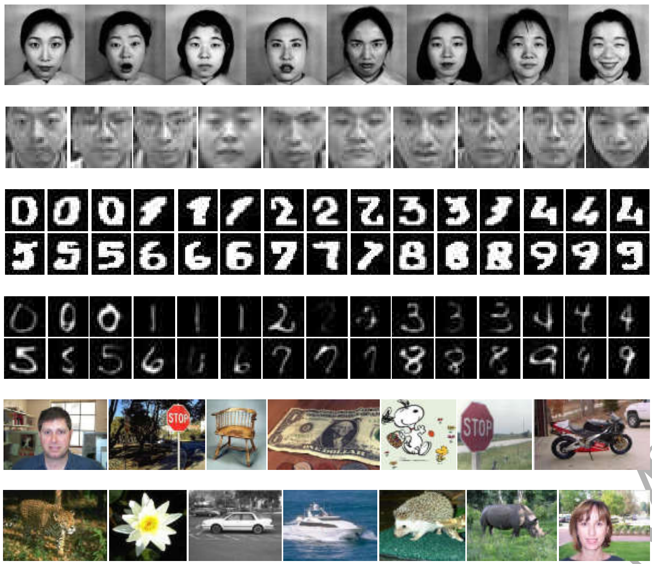


Fig. 4: Examples of six image datasets. Samples in the same row are from the same dataset: (a) JAFFE, (b) MSRA50, (c) Handwritten, (d) USPS, (e) Caltech101-7, (f) Caltech101-20.

$\#iterations \times O(Nnc + nc^2 + c^3)$. For large-scale image clustering, the number of clusters c is usually smaller than the image feature dimension n, and the number of samples N is much larger than n and c. According to the mentioned, we can conclude that the proposed algorithm requires $O(Nnc)$, which is relatively satisfactory because the computational cost is linear with respect to the number of samples N and the feature dimension n.

4. Experiment

In this section, we conduct extensive experiments on synthetic and real image datasets to illustrate the superiority of diverse fuzzy c-means. We also analyze the convergence and sensitivity of the proposed Algorithm 1 in practice.

4.1. Demonstration on Synthetic Data

In this part, we carry out experiments over two synthetic datasets to study the effectiveness of the proposed diversity regularization. The synthetic datasets are generated as follows:

- **2D synthetic data**: As shown in Figure 2a, 2000 data points are sampled from a Gaussian distribution with the mean $[10, 9]$, and a standard deviation $[5, 0; 0, 5]$. Another 50 data points are sampled from another Gaussian distribution with mean $[-3, 7]$, and a standard deviation $[2, 0; 0, 2]$.
- **3D synthetic data**: As shown in Figure 3a, three Gaussian distributions are assigned with the mean of $[0, 0, 20]$, $[0, 40, -30]$ and $[40, 0, -30]$ respectively. We separately sample $3,000$, 200, and 100 data points from the mentioned three Gaussian distributions.

In Figure 2 and Figure 3, we demonstrate the clustering results by utilizing c-means, fuzzy c-means and our diverse fuzzy c-means over two synthetic datasets. We can conclude that the traditional c-means and fuzzy c-means algorithms pay more attention to the larger cluster and fail to identify the smaller clusters actually. Similar cluster centers are obtained under this two methods because they neglect the relations between different cluster centers. The proposed diverse fuzzy c-means performs much better than other methods, which owes the diversity term to learn clustering centers with spatial completeness.

4.2. Demonstration on Real Data

4.2.1. Dataset Description

Six benchmark image datasets are used to evaluate the performance of diverse fuzzy c-means, including JAFFE, MSRA50, Handwritten, USPS, Caltech101-7, and Caltech101-20. Some examples of these datasets are shown in Figure 4. The datasets are detailed as follows:

- **JAFFE** (Lyons et al., 1998): The Japanese Female Facial Expression database (JAFFE) contains 213 gray-scale images posed by 10 Japanese females. There are seven facial expressions in the database, including happy, angry, disgust, fear, sad, surprise and neutral (Chang et al., 2016b). Each face image is represented as 576-dimensional gray pixel values.
- **MSRA50** (Liu et al., 2011): The MSRA50 is a benchmark face dataset, which consists of 1,799 frontal face images from 12 individuals. In the experiment, each image is resized to 32×32 and we directly use the 1024-dimensional pixel values to represent the face images.

Table 1: Performance comparison in terms of ACC ± STD for different methods over the six datasets.

Dataset	JAFFE	MSRA50	Handwritten	USPS	Caltech101-7	Caltech101-20
c-Means	0.798±0.042	0.472±0.045	0.632±0.035	0.617±0.038	0.501±0.021	0.481±0.032
FCM	0.942±0.021	0.538±0.032	0.808±0.024	0.667±0.031	0.799±0.025	0.592±0.031
KFCM (Polynomials)	0.952±0.032	0.595±0.028	0.832±0.028	0.669±0.028	0.800±0.029	0.597±0.017
KFCM (Radial)	0.951±0.037	0.599±0.032	0.835±0.035	0.676±0.025	0.810±0.025	0.602±0.024
RDFCM	0.957±0.023	0.602±0.041	0.825±0.025	0.711±0.035	0.809±0.025	0.572±0.026
Our Method	**0.971±0.031**	**0.645±0.026**	**0.862±0.027**	**0.741±0.032**	**0.874±0.035**	**0.687±0.018**

Table 2: Performance comparison in terms of NMI ± STD for different methods over the six datasets.

Dataset	JAFFE	MSRA50	Handwritten	USPS	Caltech101-7	Caltech101-20
c-Means	0.840±0.041	0.525±0.036	0.690±0.029	0.597±0.033	0.515±0.041	0.549±0.036
FCM	0.936±0.034	0.602±0.024	0.746±0.035	0.625±0.041	0.626±0.031	0.591±0.026
KFCM (Polynomials)	0.940±0.025	0.652±0.035	0.765±0.016	0.635±0.015	0.632±0.031	0.601±0.016
KFCM (Radial)	0.941±0.028	0.644±0.029	0.759±0.028	0.637±0.028	0.628±0.024	0.598±0.028
RDFCM	0.943±0.024	0.632±0.031	0.751±0.036	0.658±0.026	0.631±0.036	0.610±0.041
Our Method	**0.956±0.021**	**0.695±0.029**	**0.785±0.024**	**0.689±0.014**	**0.687±0.019**	**0.671±0.028**

Table 3: Performance comparison in terms of ARI ± STD for different methods over the six datasets.

Dataset	JAFFE	MSRA50	Handwritten	USPS	Caltech101-7	Caltech101-20
c-Means	0.717±0.032	0.296±0.041	0.571±0.036	0.504±0.032	0.412±0.024	0.390±0.036
FCM	0.910±0.032	0.362±0.041	0.682±0.035	0.548±0.028	0.705±0.041	0.420±0.020
KFCM (Polynomials)	0.920±0.021	0.436±0.032	0.702±0.018	0.568±0.033	0.722±0.026	0.495±0.032
KFCM (Radial)	0.920±0.028	0.441±0.027	0.712±0.020	0.567±0.028	0.723±0.036	0.510±0.016
RDFCM	0.922±0.031	0.451±0.031	0.698±0.026	0.578±0.019	0.725±0.021	0.526±0.019
Our Method	**0.940±0.024**	**0.501±0.024**	**0.741±0.031**	**0.645±0.026**	**0.786±0.023**	**0.672±0.030**

- **Handwritten** (Frank and Asuncion, 2010): Handwritten is a benchmark handwritten digit dataset of 0 to 9. It consists of 2,000 images with 200 in each category. We utilize the published feature, *i.e.* 240-dimension pixel averages in 2×3 windows (Pix), as the image representations.
- **USPS** (Hull, 1994): The US Postal handwritten digits database (USPS) contains 9,298 gray-scale handwritten digit images among 10 categories from 0 to 9 (Chang et al., 2016a). Each image is resized to 16×16 and the 256-dimensional pixel values are used to represent images.
- **Caltech101-7 and Caltech101-20** (Dueck and Frey, 2007): Caltech101-7 and Caltech101-20 are the subsets of image dataset Caltech101 (Fei-Fei et al., 2007; Chang et al., 2017d) which contains 101 common object categories. Caltech101-7 consists of 1,474 images among 7 widely used classes and Caltech101-20 contains 2,386 images selected from 20 categories. Each image is represented as the 1,984-dimensional Histogram of Oriented Gradient (HOG) feature (Dalal and Triggs, 2005).

4.2.2. Baseline Algorithms

To extensively evaluate the effectiveness of diverse fuzzy *c*-means, we compare the proposed method with the following four clustering algorithms:

- ***c*-Means** (Hartigan and Wong, 1979): *c*-Means is a popular hard *c*-partition clustering algorithm. It aims to partition all data points into *c* clusters in which each point belongs to the cluster with the nearest distance.
- **FCM** (Dunn, 1973; Bezdek, 2013): Fuzzy *c*-means (FCM) allows one image data to belong to two or more clusters, which is one of the most widely used fuzzy clustering algorithms.
- **FKCM** (Wu et al., 2003): Fuzzy kernel *c*-means (KFCM) integrates Mercer kernel function into traditional fuzzy *c*-means. In this experiment, two typical kernel functions, *i.e.* Polynomials and Radial basis functions, are used to compare with our method.
- **RDFCM** (Zhang et al., 2017): Reconstruction data fuzzy *c*-means (RDFCM) introduces the reconstructed data supervised by the original data into traditional fuzzy *c*-means. There are three variables to be optimized, including cluster centers, memberships, and reconstructed data.

4.2.3. Performance Comparison

For a fair comparison, we utilize the best tuned to record the clustering results for all competitors and our method. The

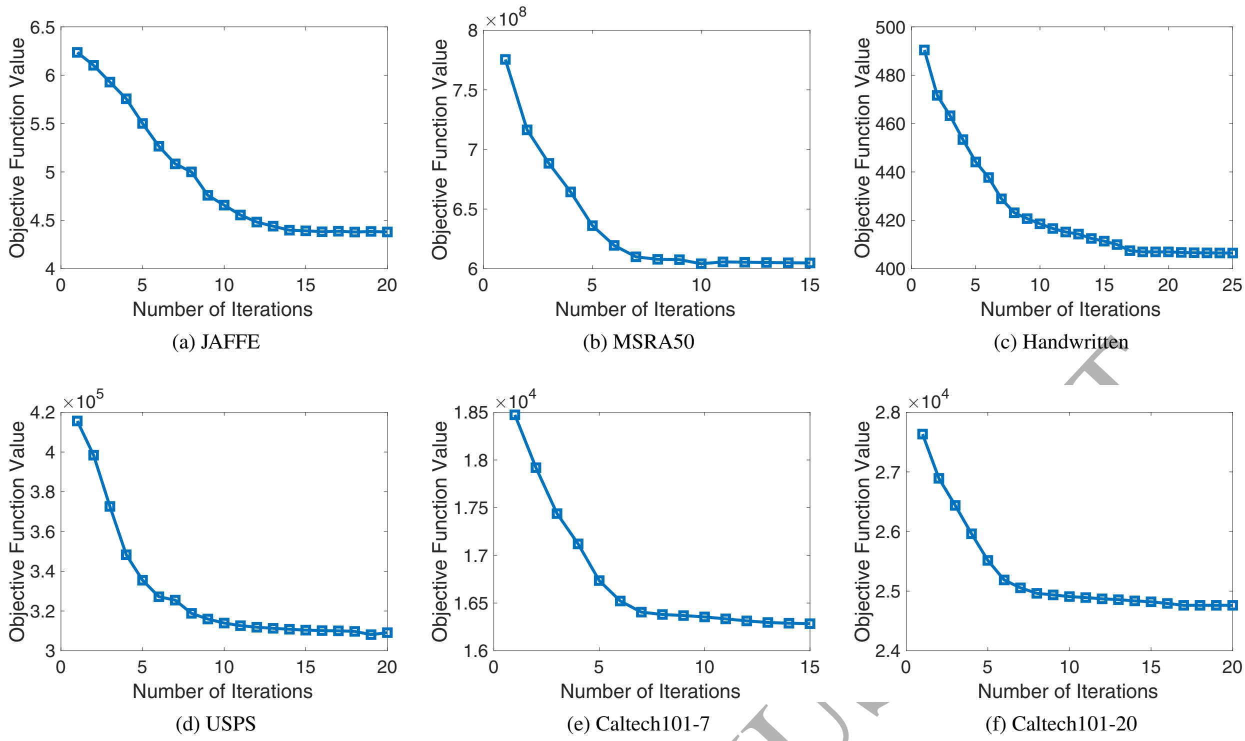


Fig. 5: Convergence curves for the value of objective function over six datasets.

c-means clustering algorithm is parameter-free. The fuzzy index parameter in FCM, KFCM, RDFCM and our method is assigned varying from 1.1 to 1.6 with step-size 0.1. For KFCM, two parameters including the base bias and exponent in polynomial kernel function are tuned in $[1.1, 2.2]$, and the parameter of standard deviation in radial kernel function is assigned in $[0.2, 3]$. For RDFCM, the gain factor for reconstruct data is set varying from 0.1 to 1 with step-size 0.1. Finally, the parameters λ and ρ in our method are tuned in $\{10^{-5}, 10^{-4}, \cdots, 10^{0}\}$. Following related studies on image clustering, we adopt three popular metrics, *i.e.* clustering accuracy (ACC), normalized mutual information (NMI) (Cai and Chen, 2015) and adjusted rand index (ARI) (Fahad et al., 2014) to measure the performance of clustering algorithms. The larger values of ACC, NMI and ARI indicate a better clustering performance. For each method, we repeat the experiment 30 times to compute the average and standard deviation (STD) on these three metrics.

In this case, we report the clustering results over six image datasets from Table 1 to Table 3 and conclude the following two conclusions: (1) Compared with five baselines, the proposed method consistently performs better over all datasets on three metrics. It indicates that the diversity regularization is effective to improve the clustering performance by controlling the learned centers to be diverse in the image data space. (2) The typical hard c-partition method, namely c-means, apparently performs poorer than other soft c-partition methods. This is because the soft c-partition methods relax the strict requirement that each image must belong to a single cluster exclusively.

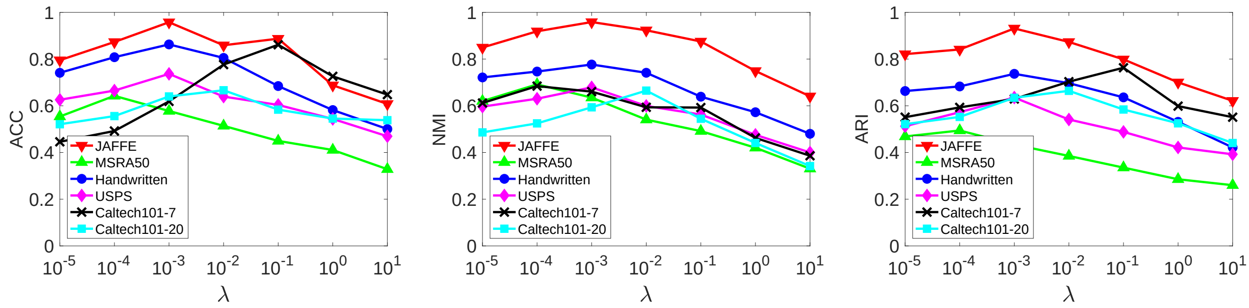
4.3. Convergence Analysis

To solve the objective function (5) for diverse fuzzy c-means, we propose the alternating optimization Algorithm 1 and demonstrate its convergence in theory by Theorem 3. In this section, we conduct numeric experiments over six image datasets to evaluate the convergence speed of the optimization problem in practice. We assign the value of hyper-parameters $\lambda = 0.01$, $m = 1.4$ and $\rho = 0.001$ and then plot the convergence curves of the proposed algorithm over all datasets in Figure 5. We can see that the objective value converges within 30 iterations for all datasets, particularly the datasets MSRA50 and Caltech101-7 just require less than 15 iterations. Therefore, the proposed algorithm is effective to achieve fast convergence.

4.4. Sensitivity Analysis

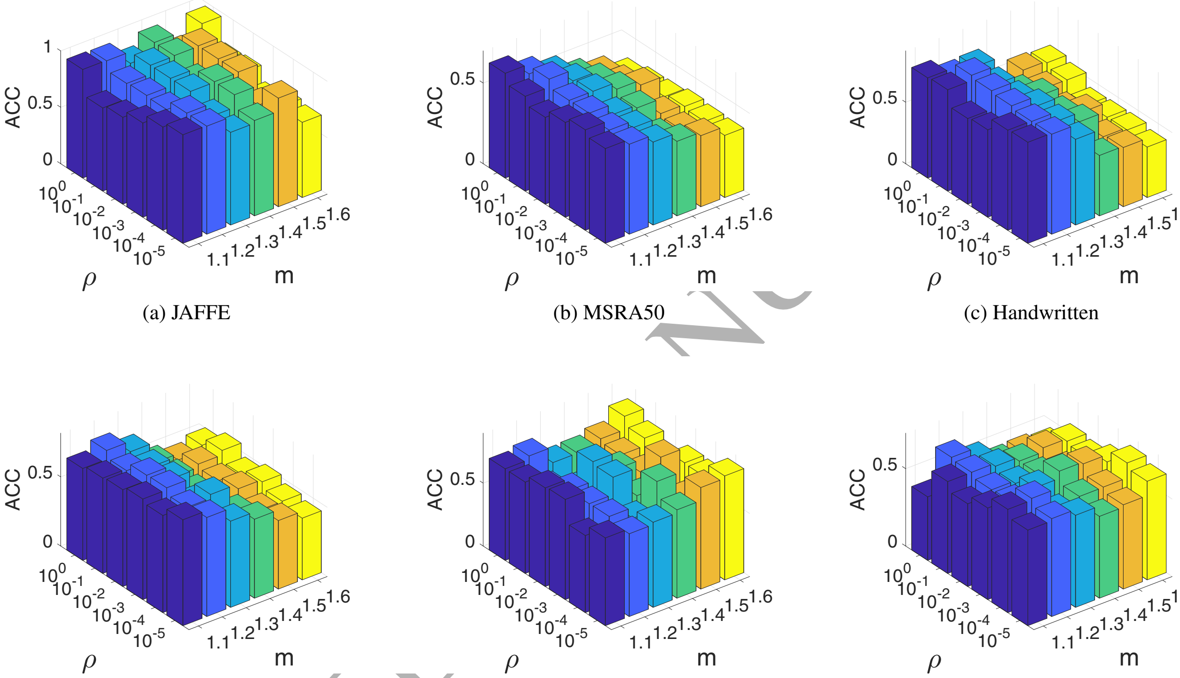
There are three hyper-parameters in the proposed Algorithm 1, including the regularization parameter λ, the fuzzy index m, and the step-size ρ for updating $\widetilde{V}$. In this section, we do some experiments to analyze the influences of these hyper-parameters on clustering performance over six image datasets.

With fixed value $m = 1.2$ and $\rho = 0.001$, we plot sensitivity curves for regularization parameter λ in Figure 6, where λ is tuned in $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}\}$. The results indicate that clustering performance on ACC, NMI and ARI over all image datasets has the analogous variation trend with the increase of λ. To be specific, the ACC, NMI and ARI values improve gradually with the increasing value of λ until the maximum is achieved; After that, these values begin to decrease when the λ is still increasing. A proper value of λ could achieve the best clustering performance by introducing



(a) ACC (b) NMI (c) ARI

Fig. 6: Sensitivity analysis on parameter λ in terms of ACC, NMI and ARI over six datasets.



(a) JAFFE (b) MSRA50 (c) Handwritten

(d) USPS (e) Caltech101-7 (f) Caltech101-20

Fig. 7: Sensitivity analysis on parameters ρ and m in terms of ACC over six datasets.

the diversity constraint into traditional fuzzy c-means. We also evaluate the sensitivity of proposed algorithm on parameters m and ρ, where m and ρ are respectively tuned in $[1.1, 1.6]$ and $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}\}$. With $\lambda = 0.1$, we plot the sensitivity column charts for six image datasets on metric ACC in Figure 7. The results indicate that the selection of m and ρ has impacts on the clustering performance, where the optimal parameters over different datasets are distinguishable.

According to the mentioned above, the image clustering performance of diverse fuzzy c-means is fluctuant with varying values of parameters λ, ρ and m. Overall speaking, when these parameters are respectively tuned in intervals $[10^{-4}, 10^{-2}]$, $[10^{-3}, 10^{-2}]$ and $[1.2, 1.3]$, the clustering performance could be satisfactory and relatively stable. The best values of these three parameters also could be obtained by cross-validation strategy.

5. Conclusion

The least-square based fuzzy c-means often suffers from the "cluster one-sidedness" problem for image clustering. To combat this problem, we define a diversity metric for image cluster centers and propose the diverse fuzzy c-means via integrating the diversity regularization of image cluster centers into the traditional fuzzy c-means objective. Since the image cluster centers are learned to differ from each others, they will cover the image data space as much as possible, rather than is controlled by part of image data. In the framework of diverse fuzzy c-means, there is always a limitation on the cluster number which should be determined in advance. To combat this issue, we attempt to leverage our previous works on sparse representation of fuzzy systems (Luo et al., 2013, 2014) and integrate the determination of cluster number into the process of fuzzy partition with diversity regularization. Additionally, we also want to apply the diverse fuzzy c-means on event detection (Chang et al., 2017c) or other applications (Chang et al., 2017b; Chang and Yang, 2017).

Acknowledgments

This work was supported by National Key Research and Development Program of China (2016YFB1000903), National Natural Science Foundation of China (61532004, 61532015, 61672418 and 61672419), Innovative Research Group of the National Natural Science Foundation of China(61721002), Innovation Research Team of Ministry of Education (IRT_17R86), Project of China Knowledge Centre for Engineering Science and Technology.

References

Bezdek, J.C., 2013. Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media.

Bhat, A., 2014. Possibility fuzzy c-means clustering for expression invariant face recognition. International Journal on Cybernetics & Informatics 3.

Boyd, S., Vandenberghe, L., 2004. Convex optimization. Cambridge University Press.

Cai, D., Chen, X., 2015. Large scale spectral clustering via landmark-based sparse representation. IEEE Transactions on Cybernetics 45, 1669–1680.

Chang, J., Wang, L., Meng, G., Xiang, S., Pan, C., 2017a. Deep adaptive image clustering, in: Computer Vision and Pattern Recognition, pp. 5879–5887.

Chang, X., Ma, Z., Lin, M., Yang, Y., Hauptmann, A.G., 2017b. Feature interaction augmented sparse learning for fast kinect motion detection. IEEE Transactions on Image Processing 26, 3911–3920.

Chang, X., Ma, Z., Yang, Y., Zeng, Z., Hauptmann, A.G., 2017c. Bi-level semantic representation analysis for multimedia event detection. IEEE Transactions on Cybernetics 47, 1180–1197.

Chang, X., Nie, F., Wang, S., Yang, Y., Zhou, X., Zhang, C., 2016a. Compound rank-k projections for bilinear analysis. IEEE Transactions on Neural Networks and Learning Systems 27, 1502–1513.

Chang, X., Nie, F., Yang, Y., Zhang, C., Huang, H., 2016b. Convex sparse pca for unsupervised feature learning. ACM Transactions on Knowledge Discovery from Data 11, 3.

Chang, X., Yang, Y., 2017. Semi-supervised feature analysis by mining correlations among multiple tasks. IEEE Transactions on Neural Networks and Learning Systems 28, 2294–2305.

Chang, X., Yu, Y.L., Yang, Y., 2017d. Robust top-k multiclass svm for visual category recognition, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM. pp. 75–83.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, pp. 886–893.

Ding, J., Ma, R., Hu, X., Yang, J., Chen, S., 2010. Fuzzy c-means revisited: towards a cluster-center-free reformulation, in: Chinese Conference on Pattern Recognition, pp. 1–5.

Dueck, D., Frey, B.J., 2007. Non-metric affinity propagation for unsupervised image categorization, in: International Conference on Computer Vision, pp. 1–8.

Dunn, J.C., 1973. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. Journal of Cybernetics 3.

Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A.Y., Foufou, S., Bouras, A., 2014. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. IEEE Transactions on Emerging Topics in Computing 2, 267–279.

Fei-Fei, L., Fergus, R., Perona, P., 2007. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. Computer Vision and Image Understanding 106, 59–70.

Frank, A., Asuncion, A., 2010. Uci machine learning repository [http://archive. ics. uci. edu/ml]. irvine, ca: University of california. School of Information and Computer Science 213.

Gordon, S., Greenspan, H., Goldberger, J., 2003. Applying the information bottleneck principle to unsupervised clustering of discrete and continuous image representations, in: International Conference on Computer Vision, p. 370.

Hartigan, J.A., Wong, M.A., 1979. Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 100–108.

Hathaway, R.J., Hu, Y., 2009. Density-weighted fuzzy c-means clustering. IEEE Transactions on Fuzzy systems 17, 243–252.

Hull, J.J., 1994. A database for handwritten text recognition research. IEEE Transactions on Pattern Analysis and Machine Intelligence 16, 550–554.

Jia, J., Yu, N., Hua, X.S., 2008. Annotating personal albums via web mining, in: International Conference on Multimedia, pp. 459–468.

Kennedy, L.S., Naaman, M., 2008. Generating diverse and representative image search results for landmarks, in: International Conference on World Wide Web, pp. 297–306.

Lai, H.P., Visani, M., Boucher, A., Ogier, J.M., 2014. A new interactive semi-supervised clustering model for large image database indexing. Pattern Recognition Letters 37, 94–106.

Li, Z., Nie, F., Chang, X., Yang, Y., 2017. Beyond trace ratio: weighted harmonic mean of trace ratios for multiclass discriminant analysis. IEEE Transactions on Knowledge and Data Engineering PP, 1–1.

Liu, H., Sun, F., Fang, B., Lu, S., 2018a. Multi-modal measurements fusion for surface material categorization. IEEE Transactions on Instrumentation and Measurement 67, 246–256.

Liu, H., Wu, Y., Sun, F., Fang, B., Guo, D., 2018b. Weakly-paired multi-modal fusion for object recognition. IEEE Transactions on Automation Science and Engineering 15, 784–795.

Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y., 2011. Learning to detect a salient object. IEEE Transactions on Pattern Analysis and Machine Intelligence 33, 353–367.

Luo, M., Sun, F., Liu, H., 2013. Hierarchical structured sparse representation for t–s fuzzy systems identification. IEEE Transaction Fuzzy System 21, 1032–1043.

Luo, M., Sun, F., Liu, H., 2014. Joint block structure sparse representation for multi-input–multi-output (mimo) t–s fuzzy system identification. IEEE Transaction Fuzzy System 22, 1387–1400.

Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J., 1998. Coding facial expressions with gabor wavelets, in: International Conference on Automatic Face and Gesture Recognition, pp. 200–205.

Wang, Y., Chen, L., Mei, J.P., 2014. Stochastic gradient descent based fuzzy clustering for large data, in: International Conference on Fuzzy Systems, pp. 2511–2518.

Wu, Z.d., Xie, W.x., Yu, J.p., 2003. Fuzzy c-means clustering algorithm based on kernel method, in: International Conference on Computational Intelligence and Multimedia Applications, pp. 49–54.

Xie, P., 2015. Learning compact and effective distance metrics with diversity regularization, in: Machine Learning and Knowledge Discovery in Databases, pp. 610–624.

Xie, P., Deng, Y., Xing, E., 2015. Diversifying restricted boltzmann machine for document modeling, in: International Conference on Knowledge Discovery and Data Mining, pp. 1315–1324.

Yang, M.S., Nataliani, Y., 2017. Robust-learning fuzzy c-means clustering algorithm with unknown number of clusters. Pattern Recognition 71, 45–59.

Yang, Y., Xu, D., Nie, F., Yan, S., Zhuang, Y., 2010. Image clustering using local discriminant models and global integration. IEEE Transactions on Image Processing 19, 2761–2773.

Yuan, Y., Yuan, H., Wang, Z., Guo, L., Hongjiu, Y., 2017. Optimal control for networked control systems with disturbances: a delta operator approach. IEEE Transactions on Instrumentation and Measurement 11, 1325–1332.

Zhang, L., Zhong, W., Zhong, C., Lu, W., Liu, X., Pedrycz, W., 2017. Fuzzy c-means clustering based on dual expression between cluster prototypes and reconstructed data. International Journal of Approximate Reasoning 90, 389–410.

Zhu, L., Chung, F.L., Wang, S., 2009. Generalized fuzzy c-means clustering algorithm with improved fuzzy partitions. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 39, 578–591.