Pattern Recognition Letters 36 (2014) 29–35

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

# Lessons to learn from a mistaken optimization

László Szilágyi *

*Budapest University of Technology and Economics, Department of Control Engineering and Information Technology, Magyar tudósok krt. 2, H-1117 Budapest, Hungary*
*Sapientia University of Transylvania, Faculty of Technical and Human Sciences, Şoseaua Sighişoarei 1/C, 540485 Tîrgu Mureş, Romania*

ARTICLE INFO

*Article history:*
Received 19 February 2013
Available online 6 September 2013

Communicated by C. Luengo

*Keywords:*
Image segmentation
Fuzzy *c*-means clustering
Fuzzy local information *c*-means clustering
Suppressed fuzzy *c*-means clustering
Grouped coordinate minimization

ABSTRACT

The fuzzy local information *c*-means (FLICM) algorithm, introduced by Krinidis and Chatzis (2010), was designed to perform highly accurate segmentation of images contaminated with high-frequency noise. This algorithm includes an extra additive term to the objective function of the fuzzy *c*-means (FCM), called local descriptor fuzzy factor, allowing the labeling of a pixel to be influenced by its neighbors, thus achieving a filtering effect. Further on, the authors of FLICM claim that their algorithm does not depend on any trade-off parameter, which were present in all previous similar approaches. In this paper we investigate the theoretical foundation of FLICM and reveal some critical issues. First of all, we show that the iterative optimization algorithm proposed for the minimization of the FLICM objective function is not suitable for the given problem, it does not minimize the objective function. Instead of that, FLICM computes an FCM-like partition using distorted distances, according to the local context of each pixel, thus performing a job that is similar to the so-called suppressed fuzzy *c*-means algorithm existing in the literature. Finally we reveal the presence of a possible trade-off in the definition of the local descriptor fuzzy term, and the necessity of another factor to compensate against the size of the considered neighborhood. Such algorithms can be effective in certain scenarios, which were documented by the authors, but a deep investigation of the limitations would be beneficial.



## 1. Introduction

Generally, the aim of image segmentation is to distinguish objects from the background and different objects from each other. The fuzzy *c*-means (FCM) algorithm (Bezdek, 1981) is a very popular tool in image segmentation because it is simple to comprehend and implement, but it frequently leads to unsatisfactory results, especially when it is employed to classify the pixels according their intensity and the image is noisy. It is a well-known fact, that an FCM algorithm employed to classify the pixel intensities, represents a clustering based on global information, because the position of pixels does not influence its fuzzy memberships with respect to various clusters. During the last two decades, a long series of modified FCM algorithms were introduced (Pham and Prince, 1999; Pham, 2001; Ahmed et al., 2002; Pham, 2003; Szilágyi et al., 2003; Chen and Zhang, 2004), to improve the accuracy of classification in the presence of various types of noise. These algorithms attempted to involve local information into the objective function, allowing the categorization of each pixel to be influenced by its neighborhood, by its local context. Since the introduction of the fast and robust solution of FGFCM (Cai et al., 2007), the trend of publishing locally improved robust versions of FCM has visibly intensified, leading to another long series of recent solutions (Liao et al., 2008, Wang et al., 2008, Wang et al., 2009, Despotovic et al., 2010, Di Martino et al., 2010, Huang and Zhang, 2010, Li and Shen, 2010, Ji et al., 2011, Li et al., 2011, Chattopadhyay et al., 2011, Liu and Pham, 2012). These algorithms have been successfully integrated into applications in various domains: remote sensing (Fan et al., 2009; Mishra et al., 2012; Gong et al., 2013b; Kalaivani et al., 2013), analysis of dermoscopy images (Zhou et al., 2009), color reduction (Schaefer and Zhou, 2009), brain imaging (Caldairou et al., 2011; Liao and Zhang, 2011), geophysics (Paasche et al., 2010), sea surface temperature monitoring (Nascimento and Franco, 2009), and arrhythmia classification (Ceylan et al., 2009).

Recently, Krinidis and Chatzis (2010) introduced the so-called fuzzy local information *c*-means (FLICM) algorithm, which included an additive term into the objective function of FCM, designed to suppress high frequency noises from the image, improving the robustness of the segmentation. As reported by its authors, FLICM seems having accomplished its mission, it has reduced the number of misclassifications in certain image segmentation scenarios. However, a deeper insight into the problem can reveal some critical issues concerning FLICM. The aim of this paper is to analyze the theoretical foundation and functionality of FLICM, within the bound of the theory of optimal systems.

* Address: Sapientia University of Transylvania, Faculty of Technical and Human Sciences, Şoseaua Sighişoarei 1/C, 540485 Tîrgu Mureş, Romania. Tel.: +40 265 208 170; fax: +40 265 206 211.

*E-mail address:* lalo@ms.sapientia.ro

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.08.027

The rest of this paper is structured as follows: Section 2 takes into account the most important works related to this study. Section 3 gives an alternative optimization scheme of the FLICM objective function, using grouped coordinate minimization. Section 4 investigates the FLICM algorithm in its originally defined context. Section 5 is a short review of algorithms built upon FLICM. The last section concludes this study.

## 2. Background works

The fuzzy $c$-means clustering and the algorithms derived from it, in their original introduction, use a wide variety of notations. In this study, they are all translated to one language. Throughout this whole paper, the partition matrix $U$ has two indices, the first one refers to the cluster, while the second one to the identity of the input vector. Thus, fuzzy membership function $u_{ik}$ describes the degree to which vector (or pixel) with index $k \in \{1, 2, \ldots n\}$ belongs to cluster with index $i \in \{1, 2, \ldots c\}$. Indices $i$ and $j$ always refer to clusters, while vectors (pixels) are addressed by $k$ and $l$.

### *2.1. Fuzzy c-means clustering*

The FCM algorithm optimally partitions a set of object data $X = \{x_1, x_2, \ldots, x_n\}$ into a previously set number of $c$ clusters, solving the following minimization problem:

$$\min_{(U,V)} \left\{ J_m^{(\mathrm{FCM})}(U, V : X) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^m \|x_k - v_i\|_A^2 \right\}, \quad (1)$$

where

- $U = \{u_{ik}\},\ i = 1 \ldots c,\ k = 1 \ldots n$, represents a fuzzy $c$-partition matrix constrained by the probabilistic conditions $0 \leqslant u_{ik} \leqslant 1$, and $\sum_{i=1}^{c} u_{ik} = 1\ \forall i = 1 \ldots c$;
- $V = (v_1, v_2, \ldots v_c)$ is the vector of unknown cluster centers (also referred to as prototypes);
- $m$ is the fuzzy exponent, the only parameter of the algorithm, constrained by $m > 1$;
- $\|s\|_A = \sqrt{s^T A s}$ is the inner product norm induced by the positive definite, symmetrical matrix $A$.

The minimization of $J_m^{(\mathrm{FCM})}$, as presented in Bezdek (1981) and Bezdek et al. (1987) is performed within the framework of grouped coordinate minimization, which is a general iterative method. Within this framework, each iteration is separated into two half problems, or half steps (Pal et al., 1996):

$$U^{(t+1)} = \arg\min_{\mathcal{U}} \{J_m^{(\mathrm{FCM})}(\mathcal{U}, V^{(t)})\} \quad (2)$$

and

$$V^{(t+1)} = \arg\min_{\mathcal{V}} \{J_m^{(\mathrm{FCM})}(U^{(t+1)}, \mathcal{V})\}. \quad (3)$$

The above half steps are called explicit ones, or analytically exact half steps, if there exist explicit formulas $U^{(t+1)} = \Psi_U(V^{(t)})$ and $V^{(t+1)} = \Psi_V(U^{(t+1)})$ for every half iterate. Such formulas in case of the FCM algorithm are deduced from zero gradient conditions of the objective function, using Lagrange multipliers.

In such optimization problems, there are also cases, when one of the half steps or both of them can only be defined implicitly. These cases require numerical optimization, usually solved via Newton's method.

The zero gradient conditions and Lagrange multipliers stand at the foundation of the theorem of FCM clustering, formulated by Bezdek (1981), which states that $(U, V)$ may minimize $J_m^{(\mathrm{FCM})}$ if and only if:

$$U^{(t+1)} = \Psi_U(V^{(t)}, X) = \{u_{ik}^{(t+1)}\} \quad \text{with} \quad (4)$$

$$u_{ik}^{(t+1)} = \frac{\|x_k - v_i^{(t)}\|_A^{-2/(m-1)}}{\sum_{j=1}^{c} \|x_k - v_j^{(t)}\|_A^{-2/(m-1)}} \quad \forall\ i = 1 \ldots c, \quad \forall\ k = 1 \ldots n \quad (5)$$

and

$$V^{(t+1)} = \Psi_V(U^{(t+1)}, X) = (v_i^{(t+1)}) \quad \text{with} \quad (6)$$

$$v_i^{(t+1)} = \frac{\sum_{k=1}^{n} \left(u_{ik}^{(t+1)}\right)^m x_k}{\sum_{k=1}^{n} \left(u_{ik}^{(t+1)}\right)^m} \quad \forall i = 1 \ldots c. \quad (7)$$

After adequate initialization of cluster prototype values $v_i^{(0)}$, Eqs. (5) and (7) are alternately applied until cluster prototypes stabilize, that is, $\|V^{(t+1)} - V^{(t)}\| < \varepsilon$, where $\varepsilon$ represents a previously defined small positive constant number. The expressions presented in Eqs. (5) and (7) are necessary conditions to minimize $J_m^{(\mathrm{FCM})}$. The above algorithm is called the alternative optimization (AO) scheme of FCM.

If we wish to defuzzify the solution, each object vector $x_k$ will be assigned to the cluster with index $w_k$ where

$$w_k = \arg\max_i \{u_{ik}^{(\mathrm{final})},\ i = 1 \ldots c\}. \quad (8)$$

### *2.2. Suppressed fuzzy c-means clustering*

The suppressed fuzzy $c$-means (s-FCM) algorithm was introduced in Fan et al. (2003), having the declared goal of reducing the execution time of FCM by improving the convergence speed, while preserving its good classification accuracy. The s-FCM algorithm is not optimal from a rigorous mathematical point of view, as it does not minimize $J_m^{(\mathrm{FCM})}$ or any other known objective function. Instead of that, it modifies the AO scheme of FCM, by inserting an extra computational step into each iteration, placed between the partition update formula (5) and prototype update formula (7). This new step deforms the partition (fuzzy membership functions) according to the following rule:

$$\mu_{ik} = \begin{cases} 1 - \alpha + \alpha u_{ik} & \text{if } i = w_k = \arg\max_j \{u_{jk},\ j = 1 \ldots c\} \\ \alpha u_{ik} & \text{otherwise} \end{cases}, \quad (9)$$

where $\mu_{ik}$ with any $i = 1 \ldots c$ and $k = 1 \ldots n$ represent the fuzzy memberships obtained after suppression. During the iterations of s-FCM, these suppressed membership values $\mu_{ik}$ will replace $u_{ik}$ in Eq. (7).

Suppression can be explained in words as follows: in each iteration, clusters compete for each input vector $x_k$, and the prototype situated closest ($v_{w_k}$) wins the competition. Fuzzy memberships of the given vector with respect to any non-winner cluster $i$ ($i \neq w_k$) is proportionally suppressed via multiplication with the previously defined value of the suppression rate $\alpha \in [0, 1]$, while the winner fuzzy membership is increased such that the modified membership values $\mu_{ik}$ still fulfil the probabilistic constraint.

In a previous paper Szilágyi et al. (2010a) we have shown, that the proportional suppression of non-winner fuzzy memberships is mathematically equivalent with a virtual reduction of the distance between the winner cluster's prototype and the given input vector. There we proved that in any iteration, for any input vector $x_k$ and its winner cluster with index $w_k$ there exists a virtually reduced distance $\delta_{w_k k} < d_{w_k k} = \|x_k - v_{w_k}\|$, for which

$$\mu_{w_k k} = \frac{\delta_{w_k k}^{-2/(m-1)}}{\delta_{w_k k}^{-2/(m-1)} + \sum_{j=1,\ j \neq w_k}^{c} d_{jk}^{-2/(m-1)}} \quad (10)$$