DENTAL IMAGE SEGMENTATION BASED ON THE COMBINATION BETWEEN FUZZY CLUSTERING AND SEMI-SUPERVISED FUZZY CLUSTERING ALGORITHMS USING SPATIAL INFORMATION

Tran Manh Tuan[1*], Tran Thi Ngan[1], Nguyen Thuy Hoa[1], Nguyen Ba Loi[2]

[1] *School of Information and Communication Technology, Thainguyen University, Vietnam*

[2] *Medical Vocational School Ba Ria – Vung Tau, Vietnam*

---***---

Abstract - ***Dental image segmentation has an important role in practical dentistry. This progress supports the dentists in diagnosing effectively some dental diseases such as cavities, resorption bone around teeth, fractured stump, periodontal disease, etc. Some recent researches, the scientists used various techniques consisting of common image processing techniques, hard clustering algorithms, fuzzy clustering algorithms and semi-supervised fuzzy clustering algorithms. These techniques were applied in dental image segmentation problem. In which, semi-supervised fuzzy clustering algorithms were quite effective in the meaning of high quality in clustering. In this paper, we present an improvement by using fuzzy C-means clustering algorithm (FCM) combined with a semi-supervised fuzzy clustering algorithm and its application in image segmentation problem. The additional information used in this semi-supervised fuzzy clustering algorithm is spatial information of dental image.***

***Key Words*:** **Fuzzy clustering, Semi-supervised fuzzy clustering, spatial information, dental image, image segmentation.**

1.INTRODUCTION

Image segmentation is the first step in image processing progress. It has an important role and effects to the accuracy of following steps in image processing [8]. Dental image segmentation also is the pivotal part in practical dentistry, especially in order to support the dentists in dental disease diagnosing [11]. In recently researches, there are two groups of image segmentation techniques including the techniques based on threshold [6] and based on clustering algorithms [12]. These methods meet the challenges in determining values of threshold parameter and defining the common boundaries of teeth samples [13]. Thus, fuzzy clustering is an effective tool to solve the problems related to the quality of clusters. The fuzzy clustering algorithms always have some additional information provided by users. This kind of information gives the orientation for the clustering step. These semi-supervised fuzzy clustering algorithms are very effective in many different fields such as image processing [4], pattern recognition, face recognition [1], risk assessment [3], bankrupt prediction [9], especially in image segmentation with color images and medical images.

Therefore, the main contribution of this paper is to propose a new semi-supervised fuzzy clustering algorithm in which there is a combination between fuzzy clustering and semi-supervised fuzzy clustering algorithm. Spatial information of dental image is considered as the additional information of image dental segmentation problem. A part from that, the evaluation of new algorithm's performance is implemented based on real data sets including 56 dental X-ray images of the patient during the period 2014-2015 at Hanoi Medical University Hospital. The paper also addresses some conclusions regarding to the use of this new algorithm.

The significance of this study is to find out a new effective algorithm in dental X-ray image segmentation. To carry out this, we propose a mathematical model for the objective function of semi-supervised fuzzy clustering algorithm in the form of optimization problem and use the additional information to improve the quality of clustering. Based on a sample set of real dental X-ray images, the model is evaluated and compared with other related algorithms. The application of this algorithm in the segmentation problem plays a significant role in following image processing progress.

The rest of the paper is organized as follows. In Section 2, we give the preliminary about the fuzzy clustering algorithms (fuzzy clustering-FCM and semi-supervised standard fuzzy clustering-SSSFC). Section 3 proposes a new semi-supervised fuzzy algorithm based on the combination between FCM and SSSFC in which additional information is the spatial information. Some experimental results are shown in Section 4 to validate the performance of new algorithm on a real dataset. Some conclusions and future works are given in Section 5.

2. Preliminary about the fuzzy clustering algorithms

In this Section, we present briefly a fuzzy clustering algorithm (FCM) in Subsection 2.1 and semi-supervised standard fuzzy clustering algorithm SSSFC in Subsection 22.2.

2.1 Fuzzy C-means algorithm

Fuzzy C-Means (FCM), proposed by Bezdek, Ehrlich & Full [2], is mainly based on the iteration process to optimize the membership matrix and the cluster centers. The objective function of FCM is defined as follows:

$$J=\sum_{k=1}^{N}\sum_{j=1}^{C}u_{kj}^{m}\left\|X_k-v_j\right\|^2\rightarrow \min \qquad (1)$$

Where m is fuzzier; C is the number of clusters; N is the number of data elements; r is the dimension of the data; u_{kj} is the membership degree of data elements X_k to cluster j; $X_k\in R^r$ is the k^{th} element of $X=\{X_1,X_2,...,X_N\}$; V_j is the center of cluster j.

Then the constraints of the objective function in (1) are:

$$\sum_{j=1}^{C}u_{kj}=1;\qquad u_{kj}\in[0,1];\forall k=\overline{1,N} \qquad (2)$$

The optimal solutions of problem (1)-(2) obtained from Lagrange multiplier method are centers of clusters in equation (3) and membership degree in equation (4).

$$v_j=\frac{\sum_{k=1}^{C}u_{kj}^{m}X_k}{\sum_{k=1}^{C}u_{kj}^{m}} \qquad (3)$$

$$u_{kj}=\frac{1}{\sum_{i=1}^{C}\left(\frac{\left\|X_k-v_j\right\|}{\left\|X_k-v_i\right\|}\right)^{\frac{1}{m-1}}} \qquad (4)$$

The FCM algorithm is presented as in Table 1

Table 1. Fuzzy C-means algorithm

Input	Dataset X including N elements in r-dimension space; The number of clusters C; fuzzier m; threshold; the largest number of iterations MaxStep
Output	Matrix U and centers of clusters V
FCM	
1	t = 0
2	$u_{kj}^{(t)}\leftarrow random;\quad \left(k=\overline{1,N};\ j=\overline{1,C}\right)$ satisfy the condition (2)
3	**Repeat**
4	t = t + 1
5	Compute $V_j^{(t)};\left(j=\overline{1,C}\right)$ by formula (3)
6	Compute $u_{kj}^{(t)};\left(j=\overline{1,C}\right)$ by formula (4)
7	**Until** $\left\|U^{(t)}-U^{(t-1)}\right\|\leq\varepsilon$ or t > MaxStep

2.2 Semi-supervised fuzzy clustering algorithm

Semi-supervised fuzzy clustering algorithms are constructed from fuzzy clustering with the additional information provided by users. Additional information makes the clustering progress have a higher quality. There are three kinds of additional information consisting of Must-link and Cannot-link constraints, labeled a part of data, pre-defined membership degree.

Yasunori et al. [15] proposed a semi-supervised fuzzy clustering with the objective function as follow

$$J=\sum_{k=1}^{N}\sum_{j=1}^{C}\left|u_{kj}-\bar{u}_{kj}\right|^{m}\left\|X_k-v_j\right\|^2\rightarrow \min \qquad (5)$$

satisficing the constraints in (2). Then the additional membership degree of data point X_k belonging to clustering C_j is denoted by $\bar{u}_{kj}$. This additional membership matrix satisfies below condition

$$\overline{U} = \left\{\bar{u}_{kj} \mid \bar{u}_{kj} \in [0,1], k = \overline{1,N}, j = \overline{1,C}\right\},$$

Based on the constraints in (2) and the objective function (5) we have

$$v_j = \frac{\sum_{k=1}^{C} \left|u_{kj} - \bar{u}_{kj}\right|^m X_k}{\sum_{k=1}^{C} \left|u_{kj} - \bar{u}_{kj}\right|^m}; \quad j = \overline{1,C} \qquad (6)$$

Where u_{kj} are defined differently in two instances as follow

- $m > 1$:

$$u_{kj} = \bar{u}_{kj} + \left(1 - \sum_{i=1}^{C} \bar{u}_{ki}\right) \frac{\left(\frac{1}{\|X_k - v_j\|}\right)^{\frac{2}{m-1}}}{\sum_{i=1}^{C} \left(\frac{1}{\|X_k - v_i\|}\right)^{\frac{2}{m-1}}}, \; k = \overline{1,N}, \; j = \overline{1,C}. \qquad (7)$$

- m=1:

$$u_{kj} = \begin{cases} \bar{u}_{kj} + 1 - \sum_{j=1}^{C} \bar{u}_{kj}, & khi\ k = \arg\min_i \|X_k - V_i\|^2 \\ \bar{u}_{kj}, & khi\ k \neq \arg\min_i \|X_k - V_i\|^2 \end{cases}, \; k = \overline{1,N},$$

$$j = \overline{1,C}. \qquad (8)$$

Semi-Supervised Standard Fuzzy Clustering algorithm (SSSFC) is presented as in Table 2

Table 2. Semi-Supervised Standard Fuzzy Clustering algorithm

Input	Datasets X includes N elements; the number of clusters C; additional membership matrix $\overline{U}$, Threshold ε; the maximum number of iterations maxStep> 0
Output	Matrix U and cluster centers V
SSSFC	
1:	t = 0
2:	Initialize randomly clustering centers $V^{(t)}$
3:	**Repeat**
4:	Calculate $U^{(t)}$ by defined values $\bar{u}_{kj}$ using equation (7) if $m > 1$ or using equation (8) in the case of m=1.
5:	t = t + 1
6:	Calculate $V^{(t)}$ by v_j using (6)
7:	**Until** $\left\|V^{(t)} - V^{(t-1)}\right\| \le \varepsilon$ or t > maxStep

3. The novel combination of fuzzy clustering and semi-supervised clustering

3.1 The main mechanism of proposed model

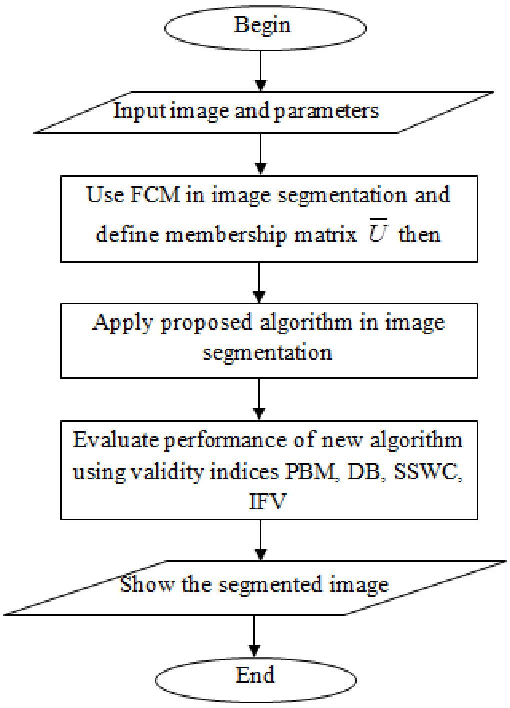


Fig. 1. The diagram of new algorithm (SSFC-DF)

In Figure 1, a novel model combining fuzzy clustering algorithm and semi-supervised fuzzy algorithm is presented. In which, additional information of semi-supervised fuzzy clustering algorithm is spatial information. Firstly, from input image, we use FCM to get the output membership matrix as the additional information of semi-supervised fuzzy clustering algorithm. After that, the SSFC-SC is applied into the original image. The clustering results are evaluated by various criteria together with the comparison among this proposed model and other related methods.

3.2. The proposed combination model (SSFC-DF)

Based on the objective function of fuzzy clustering algorithm (FCM) and semi-supervised fuzzy algorithm (SSSFC) shown in Section 2, we propose a new objective function. This combination objective function has three main components: fuzzy clustering objective function, spatial information and semi-supervised fuzzy clustering function. It can be written as in equation (9) below.

$$J = \sum_{k=1}^{N}\sum_{j=1}^{C}(u_{kj})^2\|x_k - v_j\|^2 + \sum_{k=1}^{N}\sum_{j=1}^{C}(u_{kj})^2 R_{kj}^2 + \sum_{k=1}^{N}\sum_{j=1}^{C}\left|u_{kj} - \bar{u}_{kj}\right|^2\|x_k - v_j\|^2 \to \min \quad (9)$$

This function must satisfy both constraints of fuzzy clustering algorithm and semi-supervised fuzzy algorithm:

$$\sum_{j=1}^{C} u_{kj} = 1; \qquad u_{kj} \in [0,1]; \qquad \forall k = \overline{1,N} \quad (10)$$

$$\sum_{j=1}^{C} \bar{u}_{kj} \le 1; \qquad \bar{u}_{kj} \in [0,1]; \qquad \forall k = \overline{1,N} \quad (11)$$

Where C is the number of clusters, N is the number of objects in dataset, r is the dimension of each object, u_{kj} is the membership degree of k^th^ data point belonging to j^th^ cluster; $\bar{u}_{kj}$ is the additional information (defined from output of FCM).

Denote that $X_k \in R^r$ is k^th^ point of dataset $X = \{X_1, X_2, ..., X_N\}$; v_j is the center of j^th^ cluster, R_{kj} is a function used to calculate space distance [10] from X_k to V_j.

For solving the optimal problem (9-11), the Lagrange multiplier method is used. Taking the derivative of (9) by V, we obtain:

$$\frac{\partial J}{\partial v_j} = -2\sum_{k=1}^{N}(u_{kj})^2(x_k - v_j) - 2\sum_{k=1}^{N}(u_{kj})^2\left|u_{kj} - \bar{u}_{kj}\right|(x_k - v_j) \quad (12)$$

Setting up the value of this derivation to zero, we find the cluster centers V as in (13).

$$v_j = \frac{\sum_{k=1}^{N}\left((u_{kj})^2 + \left|u_{kj} - \bar{u}_{kj}\right|\right)x_k}{\sum_{k=1}^{N}\left((u_{kj})^2 + \left|u_{kj} - \bar{u}_{kj}\right|\right)} \quad (13)$$

Then Lagrange function is defined as

$$L(U) = J + \sum_{k=1}^{N}\lambda_k\left(\sum_{j=1}^{C}u_{kj} - 1\right) \quad (14)$$

The partial derivation of L(U) by U is presented by following equation

$$\frac{\partial L}{\partial u_{kj}} = 2u_{kj}\|x_k - v_j\|^2 + 2u_{kj}R_{kj}^2 + 2u_{kj} - \bar{u}_{kj}\|x_k - v_j\|^2 + \lambda_k \quad (15)$$

The membership degree is defined by using (16):

$$u_{kj} = \frac{-\lambda_k + 2\bar{u}_{kj}\|x_k - v_j\|^2}{2*\left(2\|x_k - v_j\|^2 + R_{kj}^2\right)} \quad (16)$$

From the constraint in (10), we have

$$\lambda_k = \left(\sum_{j=1}^{C}\frac{\bar{u}_{kj}\|x_k - v_j\|^2}{\left(2\|x_k - v_j\|^2 + R_{kj}^2\right)} - 1\right) \Bigg/ \left(\sum_{j=1}^{C}\frac{1}{2\left(2\|x_k - v_j\|^2 + R_{kj}^2\right)}\right) \quad (17)$$

The optimal solutions obtained from Lagrange multiplier method are centers of clusters in (13) and membership degree in (16)-(17).

The detail steps of novel model are presented in Table 3 below

Table 3. The proposed method (SSFC-DF) in dental image segmentation

Input	Datasets X includes N elements; the number of clusters C; additional membership matrix $\bar{U}$, Threshold ε; the maximum number of iterations maxStep> 0
Output	Matrix U and cluster centers V
SSFC-DF	
1:	t = 0
2:	Initialize randomly $V^{(t)}$
3:	**Repeat**
4:	Calculate $U^{(t)}$ based on $\bar{u}_{kj}$ using (16), (17).
5:	t = t + 1
6:	Calculate $V^{(t)}$ by u_{kj} using (13)
7:	**Until** $\|V^{(t)} - V^{(t-1)}\| \le \varepsilon$ or t >maxStep

4. Experiment results

4.1. Preparing for illustration

In this Section, we show the numerical results by applying new algorithm into dental image segmentation. The dataset consists of 56 dental X-ray images obtained from Hanoi Medical University Hospital in the period of 2014-2015.

Based on this real dataset, fuzzy C-means clustering (FCM) algorithm, semi-supervised standard fuzzy clustering (SSSFC) algorithm and our proposed combination model (SSFC-DF) have been implemented in Matlab 2014 and executed on a PC VAIO laptop with Core i5 processor.

The validity indices used to evaluate the performance of new model in this paper are DB [14], SSWC [14], PBM [14], IFV[7]. These indices are also used to compare the effectiveness of our model with other related algorithm (FCM, SSSFC)

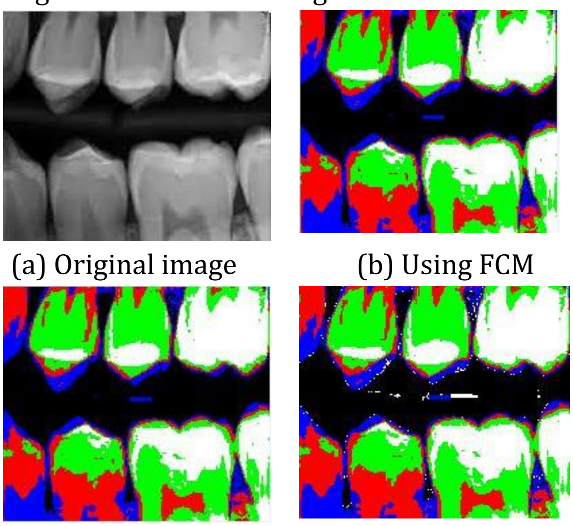
4.2. Experiment results in detail

All three algorithms are installed in the same values of parameters including the number of clusters C=10, ε = 0.005 and maxStep=150 with 20 runs. The average of values obtained after 20 runs in each index are presented as in Table 4.

Table 4. Mean of all algorithms on the given dataset (bold values is the best in a row)

	FCM	SSSFC	SSFC-SC
PBM	34590.6	87374.5	**101982.5**
SSWC	0.629	0.763	**0.874**
DB	0.658	0.562	**0.546**
IFV	30.34	70.76	**82.78**

From table 4, the performance of new model SSFC-DF has the better value than others in all of four indices PBM, SSWWC, DB, IFV. The results of dental X-ray image segment are shown in Figure 2.



(a) Original image (b) Using FCM

(c) Using SSSFC (d) Using SSFC-DF

Fig. 2. The illustration of segmentation results on Image 1

5. Conclusions

In this paper, we concentrate to propose a novel semi-supervised fuzzy clustering algorithm applied in to dental image segmentation problem. This algorithm is based on the combination between fuzzy clustering (FCM) algorithm and semi-supervised fuzzy clustering with the spatial information used as additional information. The main contributions of this paper are i) to propose a new general diagram of dental image segmentation problem (presented in Section 3.1); ii) to propose a novel semi-supervised fuzzy clustering algorithm (given in Section 3.2); iii) the experimental results on the real dataset show the evaluation and comparison of performance of the proposed algorithm with other related algorithms (presented in Section 4).

From this study, there are some ideas for the future works: i) to apply these results into typical applications such as image diagnosis support system, aesthetic dentistry support system; ii) to improve this model based on other features of dental image as additional information in semi-supervised fuzzy clustering.

REFERENCES

[1]. Agarwal, M., Agrawal, H., Jain, N., & Kumar, M. (2010). Face recognition using principle component analysis, eigenface and neural network. In *Signal Acquisition and Processing, 2010. ICSAP'10. International Conference on* (pp. 310-314). IEEE.

[2]. Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Kluwer Academic Publishers.

[3]. Chen, J., Zhao, S., & Wang, H. (2011). Risk analysis of flood disaster based on fuzzy clustering method. *Energy Procedia*, *5*, 1915-1919.

[4]. Chuang, K. S., Tzeng, H. L., Chen, S., Wu, J., & Chen, T. J. (2006). Fuzzy c-means clustering with spatial information for image segmentation. C*omputerized medical imaging and graphics*, *30*(1), 9-15.

[5]. Jain, A. K., & Chen, H. (2004). Matching of dental X-ray images for human identification. *Pattern recognition*, *37*(7), 1519-1532.

[6]. Houhou, N., Bresson, X., Szlam, A., Chan, T. F., & Thiran, J. P. (2009). Semi-supervised segmentation based on non-local continuous min-cut. In *Scale Space and Variational Methods in Computer Vision* (pp. 112-123). Springer Berlin Heidelberg.

[7]. Hu, C., Meng, L., & Shi, W. (2008). Fuzzy clustering validity for spatial data. *Geo-spatial information science*, *11*(3), 191-196.

[8]. Lim, Y. W., & Lee, S. U. (1990). On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques. *Pattern recognition*, *23*(9), 935-952.

[9]. Martin, A., Gayathri, V., Saranya, G., Gayathri, P., & Venkatesan, P. (2011). A hybrid model for bankruptcy prediction using genetic algorithm, fuzzy c-means and MARS. *arXiv preprint arXiv:1103.2110*.

[10]. Ngo, L. T., Mai, D. S., & Pedrycz, W. (2015). Semi-supervising Interval Type-2 Fuzzy C-Means clustering with spatial information for multi-spectral satellite image classification and change detection. *Computers & Geosciences*, *83*, 1-16.
[11]. Oliveira, J., & Proença, H. (2011). Caries detection in panoramic dental X-ray images. In *Computational Vision and Medical Image Processing* (pp. 175-190). Springer Netherlands.
[12]. Paiva, A. R., & Tasdizen, T. (2010). Fast semi-supervised image segmentation by novelty selection. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on* (pp. 1054-1057). IEEE.
[13]. Sujji, G. E., Lakshmi, Y. V. S., & Jiji, G. W. (2013). MRI Brain Image Segmentation based on Thresholding. *International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-3 Number-1 Issue*, 97-101.
[14]. Vendramin, L., Campello, R. J., & Hruschka, E. R. (2010). Relative clustering validity criteria: A comparative overview. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *3*(4), 209-235.
[15]. Yasunori, E., Yukihiro, H., Makito, Y., & Sadaaki, M. (2009, August). On semi-supervised fuzzy c-means clustering. In *Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on* (pp. 1119-1124). IEEE.

BIOGRAPHIES

Msc. **Tran Manh Tuan** received the Bachelor on Applied Mathematics and Informatics at Hanoi University of Science and Technology in 2003 and Master degree on Computer Science at Thainguyen University in 2007. Now, he is a reseacher at Institute of Information Technology, Academic Institute of Science and Technology, Vietnam. He worked as a lecturer at Faculty of Information Technology, School of Information and Communication Technology, Thainguyen University from 2003. Office address: University of Information and Communication Technology, Thai Nguyen University, Quyet Thang, Thai Nguyen city, Thai Nguyen, Vietnam. Email: tmtuan@ictu.edu.vn.

Tran Thi Ngan obtained the Bachelor degrees on Mathematics – Informatics at VNU University of Science, Vietnam National University (VNU). She got Master degree on Computer Science at Thai Nguyen University. She received PhD degree on applied Mathematics – Informatics at Hanoi University of Science and Technology. From 2003, she worked as a lecture in Faculty of Information Technology, School of Information and Communication Technology, Thai Nguyen University. Her major interests are discrete mathematic, Monte Carlo method, optimization, probability theory and statistics, machine learning.

Nguyen Thuy Hoa received the Bachelor on Special English at Viet Bac Teacher Training College, Thai Nguyen Unversity in 2000 and Master degree on English at Vietnam National University, Ha Noi in 2005. Now, she is a lecturer at Faculty of Basic Science of School of Information and Communication Technology, Thai Nguyen University. Her major interests are English for soft computing and medical informatics

Mr. Nguyen Ba Loi is teacher of Medical Vocational School Ba Ria – Vung Tau. Now, he is a master student at Lac Hong University. His major interests are image processing and intelligent computing.