การจำแนกภาพแบบฟังก์ชันเชิงเส้นการปรับเทียบฮีสโตแกรมผูกพันบนพื้นฐานทฤษฎีอัตสารสนเทศและขั้นตอนวิธีการจัดกลุ่มค่าเฉลี่ยกลุ่มแบบฟัซซีในการส่งสารสนเทศ

The Histogram Equalization-Obligated Linear Function-Image Classification Based on Self-Information Theory and Fuzzy C-Means Clustering Algorithm in Information Transmission

บงการ หอมนาน*
Bongkarn Homnan
วาทิต เบญจพลกุล**
Watit Benjapolakul

*นักวิจัย ศูนย์วิจัย มหาวิทยาลัยธุรกิจบัณฑิตย์
*Researcher, Research Center, Dhurakij Pundit University.
**รองศาสตราจารย์ ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
**Associate Professor, Department of Electrical Engineering, Faculty of Engineering, Chulalongkorn University.

บทคัดย่อ

การพัฒนาสารสนเทศสื่อประสมกำลังได้รับการพัฒนาอย่างต่อเนื่องโดยเฉพาะด้านเทคโนโลยีสุขภาพอิเล็กทรอนิกส์ บทความนี้นำเสนอการจำแนกภาพสมองแบบการกำทอนแม่เหล็ก (MR) เฉพาะด้วยขั้นตอนวิธีค่าเฉลี่ยกลุ่มแบบฟัซซีและอัตสารสนเทศ การจำแนก MR ได้รับการปรับเทียบโดยฟังก์ชันเชิงเส้นผูกพันจากผลลัพธ์พร้อมจำนวนจุดภาพที่เหมาะสม ผลที่ได้พบว่าประโยชน์ของการใช้ค่าเฉลี่ยกลุ่มแบบฟัซซีและอัตสารสนเทศเงื่อนไขสำหรับภาพวิธีตำแหน่งจุดภาพทางโครงสร้างแบบฟังก์ชันเชิงเส้นการปรับเทียบฮีสโตแกรมผูกพันคือวิธีที่นำเสนอสามารถให้ค่าสารสนเทศทางภาพที่ละเอียดกว่า นอกจากนั้นเอนโทรปีสารสนเทศของภาพฐานฮีสโตแกรมตอบสนองแสดงให้เห็นว่าวิธีที่นำเสนอสามารถให้อัตราส่วนการบีบอัดข้อมูลฐานเอนโทรปีสารสนเทศต่ำกว่าด้วย

คำสำคัญ: การจำแนก ฮีสโตแกรม การปรับเทียบ ฟัซซี ภาพ สารสนเทศ

Abstract

The development of the multimedia information is continually growing especially in an e-health technology. The specific Magnetic Resonance (MR) brain image classification using fuzzy c-means clustering algorithm and the self-information in this paper is proposed. Retained and with the optimum number of pixels, the classified MR has been histogram equalized by the obligated linear function. As a result, the benefit of fuzzy c-means clustering algorithm and the constrained self-information on the corresponding histogram equalization-obligated linear function-structural pixel position scheme image shows that the proposed method can give more delicate image information. In addition, the information entropy of the corresponding histogram-based image shows that the proposed method can also give the lower information entropy-based data compression ratio.

Keywords: Classification, histogram, equalization, fuzzy, image, information.

1. Introduction

The aim of this paper is to reveal the positions of the important tissues in the specific Magnetic Resonance Image (MRI). The Fuzzy C-Means Clustering Algorithm (FCMCA) or the Soft K-Means Clustering Algorithm (SKMCA) developed from the K-Means Clustering Algorithm (KMCA) is introduced to be applied. The FCMCA of an image, based on the alpha-trimmed average of the corresponding histogram extracted from the image keyframe, is suitable for the clustering. Besides the improvement of computation time of FCMCA, the FCMCA of **the constraint self-information curve** is applied in this paper to emphasize on the image information. The specific MRI based on the Structural Pixel Position Scheme (SPPS) image gives the corresponding **Histogram Equalization (HE) obligated by the linear function**, called the Histogram Equalization-Obligated Linear Function-Spatial Pixel Position Scheme (HE-OLF-SPPS: HOS) image. Compared with the HOS image, the specific MRI based on the SPPS image giving the corresponding HE, called the corresponding histogram-based image is introduced.

This paper sections include SPPS image information, FCMCA of the SPPS image, FCMCA of the self-information curve, Results and analysis, and Conclusions and future work.

2. SPPS image information

In an information theory, a self-information is a measure of the information content associated with the outcome of a random variable. The self-information $I(i)$ is as

$$I(i) = -\log(p(i)), i \epsilon \{1, 2,3, \ldots, L\} \quad (1)$$

where the function and the parameter can be denoted below:

$p(i)$ the probability of i and
L the maximum Light Intensity Level (LIL).

2.1 Corresponding histogram-based image information

In a gray-scale image, an $I(i)$ is as (1) and L can be obtained from the number of bits, n_b, of a pixel. Normally, n_b is 8 bits and L is 256 levels. Thus, the corresponding histogram of a gray-scale image can illustrate image information by each self-information of the LIL. The decreased significant number of pixels, n_p, retrieved from $I(i)$, is as

$$n_p = \sum_{i=LIL_L}^{LIL_H} n_i \quad (2)$$

where the parameters can be denoted below:

LIL_L the lower LIL and
LIL_H the higher LIL.

2.2 HOS image information

Due to the optimum consume of n_p of an HOS image, an information in the HOS image will be investigated. In this paper, the significance of $I(i)$ as (1) of the initial gray-scale image will be the constraint to retrieve the HOS image self-information. Thus, only the n_p retrieved from two constraints: the obligated linear function of the HOS image and $I(i)$, are concerned as

$$n_p = \sum_{i=LIL_L}^{LIL_H} a_1 i \quad (3)$$

where the parameter can be denoted below:

a_1 the 1st order coefficient.

3. FCMCA of the SPPS image

The FCMCA starts with the initial cluster centers of the SPPS image, placed as the center location of each cluster in an image. In addition, the FCMCA assigns each pixel value a membership grade, μ , for each cluster. Iteratively, the cluster centers update a μ for each pixel value, the FCMCA moves the cluster centers to the proper location within an image. This iteration is based on minimizing an objective function that represents the distance from each pixel to own cluster center weighted by a μ of a pixel value.

The FCMCA can partition a finite collection of n pixels $X = x_1, \dots, x_n$ into a collection of c fuzzy clusters with respect to given criterion. Given a finite set of the n pixels, the FCMCA returns a list of c cluster centers $C = c_1, \dots, c_c$ and a partition matrix $U = u_{i,j}(x)\epsilon[0,1], i = 1, \dots, n;\ j = 1, \dots, c,$ where $u_{i,j}(x)$ informs the degree to which x_i belongs to cluster c_j. Like the KMCA, the FCMCA aims to minimize an objective function $u_{i,j}(x)$ as

$$\mu_{i,j}(x) = \frac{1}{\sum_{j=1}^{c}\left(\frac{d(c_i,x)}{d(c_j,x)}\right)^{2/(\epsilon-1)}} \quad (4)$$

where the function and the parameters can be denoted below:

$d(c_i, x)$ the Euclidean distance between c_x and x,

ε the fuzzifier, and

x the $\left[p_x, p_y\right]^T$ pixel coordinate vector

The fuzzifier ε determines the level of cluster fuzziness. A larger ε results in a smaller $\mu_{i,j}$. Generally, an ε equal to 2.0 can get the common $\mu_{i,j}$. It can be noted that at ε equal to 1.0, the $\mu_{i,j}$ converges to 0.0 or 1.0, which implies a crisp partitioning.

Any point x has a set of coefficients giving the degree of being in the $w_k(x)$. With the FCMCA, the center of a cluster is the mean of all points, weighted by their degree of belonging to the cluster.

3.1 FCMCA of the corresponding histogram-based image

The FCMCA starts with the initial cluster centers of the corresponding histogram-based image, placing the center location of each cluster in the image of n_p pixels as expressed in (2). Given a finite set of the n_p pixels, the FCMCA also returns a list of c cluster centers $C = c_1, \dots, c_c$ and a partition matrix $U = u_{i,j}(x)\epsilon[0,1], i = 1, \dots, n_p;\ j = 1, \dots, c.$

3.2 FCMCA of the HOS image

The FCMCA starts with the initial cluster centers of the HOS image, placing the center location of each cluster in the image of n_p pixels as expressed in (3). Given a finite set of the n_p pixels, the FCMCA also returns a list of c cluster centers $C = c_1, \dots, c_c$ and a partition matrix $U = u_{i,j}(x)\epsilon[0,1], i = 1, \dots, n_p;\ j = 1, \dots, c.$

4. FCMCA of the self-information curve

Generally, the domain of LIL_L and LIL_H is wide enough n_p to retrieve an image information. Fortunately, a probability $p(i)$ in (1) can be computed by n_i/n. In the same way as the FCMCA of the SPPS image, the FCMCA of the self-information curve can be applied for the self-information curve of C_h clusters. The FCMCA assigns each value of $I(i)$ on the self-information curve a μ for each cluster. Iteratively, the cluster centers update a μ for each value, the FCMCA moves the cluster centers to the proper location along the self-information curve. This iteration is also based on minimizing an objective function that represents the distance from each value of n_i to its own cluster center weighted by a μ of a value of n_i. Thus, the LIL_L can be found from the domain of the corresponding histogram function, LIL, of the cluster center giving the maximum of the function range, $I(i)$. On the opposite side, the LIL_H can be normally found from LIL of the adjacent cluster center giving the sub-maximum of $I(i)$.

The FCMCA also can partition a finite collection of L LILs, $x_{LIL} = x_1, \dots, x_L$ into a collection of c_{LIL} fuzzy clusters with respect to given criterion. Given a finite set of the L LILs, the FCMCA returns a list of cluster centers $C_{LIL} = c_{1_{LIL}}, \dots, c_{c_{LIL}}$ and a partition matrix $U_{x_{LIL}} = u_{i,j}(x_{LIL}) \epsilon [0,1], i = 1, \dots, L;\ j = 1, \dots, c_{LIL},$ where $u_{i,j}(x_{LIL})$ informs the degree to which x_{LIL} belongs to cluster $c_{j_{LIL}}$. Effectively, an objective function of FCMCA of the self-information curve consisting of x_{LIL} is the same as (4) and the fuzzifier ε is set to the same value of 2.

Refer to (2) and (3), the LIL_H and LIL_L from the FCMCA of the self-information curve can be expressed, respectively, as

$$LIL_L = \left\lfloor I^{-1}\left(max\big(I(LIL)\big)\right)\right\rfloor \quad (5)$$

and

$$LIL_H = \left\lfloor I^{-1}\left(submax\big(I(LIL)\big)\right)\right\rfloor \quad (6)$$

where the function can be denoted below:

$I(LIL)$ the element in C_{LIL}.

5. Results and analysis

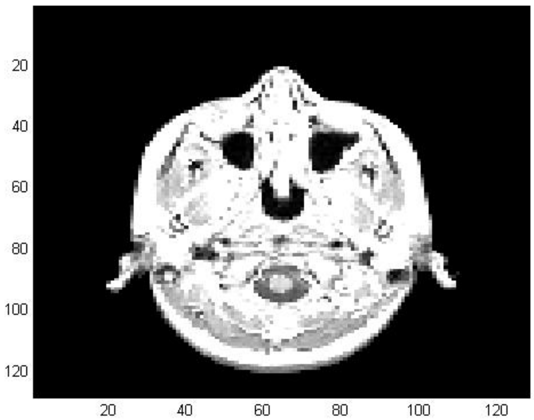
In this section, there are two subsections including the experiment results and the information entropy analysis.

5.1 Experimental results

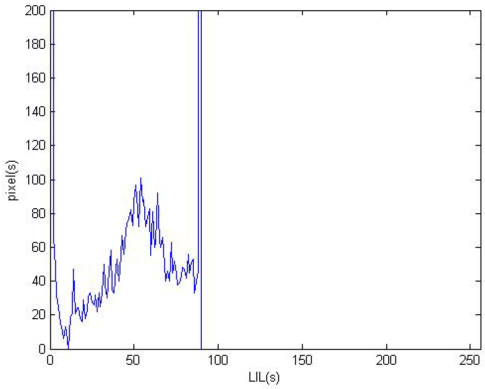
There are the input parameters including ε, a_1, C, C_h, L and n in Table 1. The original image with L of 89 and n of 16384 including the corresponding histogram can be illustrated in Fig. 1 a) and Fig. 1 b), respectively. In this experiment, there are the number of necessary holes and links of 6 in the original image. And the number of 3-points of those holes and links is equal to 18. Then, C was computed by adding 2 cluster centers to 18 cluster centers, the top of the MR brain image at the sufficient link and the reference point, to become 20 cluster centers. In addition, C_h was computed from $\log_2(89)$ = 6.476 for the 89 LIL image. The integer of 6 was reduced to 4 positions of cluster centers by neglecting 2 bound positions.

Table 1 The input parameters.

parameters	assigned values
ε	2 (Jang, J. S. R., & Sun, C. T. (1997))
a_1	0.4839 (Homnan, B., & Benjapolakul, W. (2013))
C	20
C_h	4
L	89
n	16384



(a)



(b)

Fig. 1 a) The original image and b) the corresponding histogram.

Conformed to (5), (6), and the input parameters in Table 1 with C_h = 4, the FCMCA of the self-information curve gives the integer of LIL_L and LIL_H as 10 and 33, respectively, as illustrated in Fig. 2 and Table 1

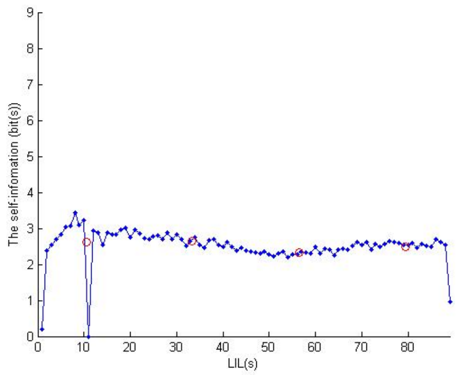


Fig. 2 The self-information of 4 cluster centers.

Table 2 The output parameters of the images.

parameters	values
LIL_L	10
LIL_H	33
n_p (The corresponding histogram-based image)	620
n_p (The HOS image)	233

Conformed to (4) with C_h = 4 and C = 20 in Table 1, the FCMCA of the self-information curve gives the obvious 19 clusters of the corresponding histogram-based image, as illustrated in Fig. 3. In addition, the FCMCA of the self-information curve gives the 20 cluster centers on the corresponding histogram-based image in the domain of $LIL_L - LIL_H$, as illustrated in Fig. 4.

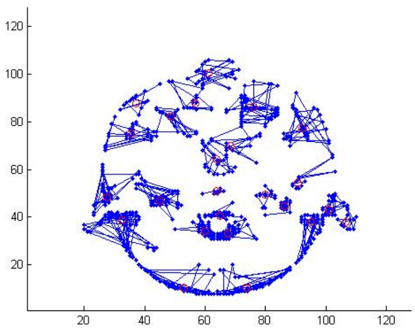


Fig. 3 The 20 cluster centers of the corresponding histogram-based image in the domain of $LIL_L - LIL_H$.

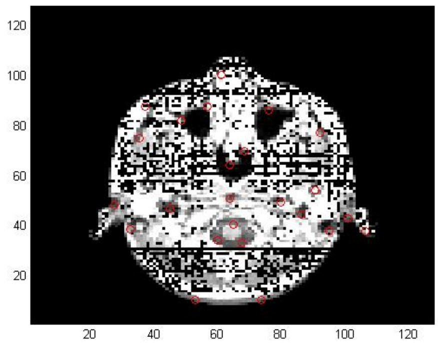


Fig. 4 The 20 cluster centers on the corresponding histogram-based image in the domain of $LIL_L - LIL_H$.

Conformed to (4) with a given a_1 = 0.4839 of the obligated linear function, C_h = 4, and C = 20, illustrated in Table 1, the FCMCA of the corresponding histogram-based image gives the 20 clusters of the HOS image as illustrated in Fig. 5. In addition, the FCMCA of the self-information curve gives the 20 cluster centers on the HOS image, as illustrated in Fig. 6.

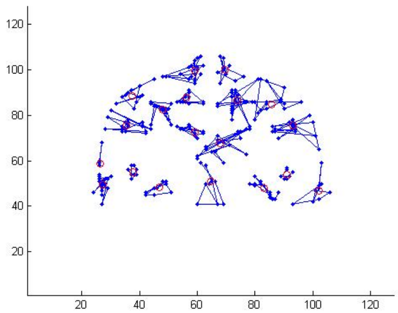


Fig. 5 The 20 cluster centers of the HOS image.

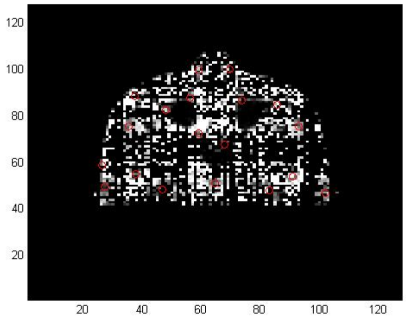


Fig. 6 The 20 cluster centers on the HOS image.

Due to the more n_p of 620 pixels, the 20 clusters can be distributed on the corresponding histogram-based image as illustrated in Fig. 3. In addition, the higher to the highest gradient portions of the corresponding histogram-based image are close to the centers of the 20 clusters. Furthermore, due to the more n_p of 233 pixels, the 20 clusters are emphasized on the upper part of the HOS image as illustrated in Fig. 5. Thus, the more important tissues of the HOS image, the specific MRI of the compressed image, especially in the upper part of the original image can be illustrated as the centers of the clusters in Fig. 5.

The n_p of the Fig. 4 and Fig. 6 can express the compression ratio, ρ, as ρ = n_p / n = 620/16384 = 0.0378 and 233/16384 = 0.0142, respectively. In the comparative analysis between the corresponding histogram-based image and the HOS image, the ratio of ρ of the HOS image in the Fig. 4 to that of the corresponding histogram-based image in Fig. 6 can be reduced to 233/620 = 0.3758 in order to obtain the important tissues of the original image. Notably, neglecting the computation times of the FCMA of the 89 point self-information curve, the FCMA of the HOS image improves the computation time to at least 37.58% compared with that of the corresponding histogram-based image, conformed to n_p.

5.2 Information entropy analysis

In an information theory, the information entropy is a measure of the amount of information in a transmitted information symbol(s) and is sometimes referred to as the Shannon entropy. The information entropy of the corresponding histogram-based image in the domain of $1 - L$, H_1^L, can be expressed in terms of a set of probabilities $p(i)$ as

$$H_1^L = -\sum_{i=1}^{L} p(i) \log p(i) \quad (7)$$

Thus, the information entropy of the corresponding histogram-based image in the domain of LIL_L - LIL_H, $H_{LIL_L}^{LIL_H}$, can be expressed as

$$H_{LIL_L}^{LIL_H} = -\sum_{i=LIL_L}^{LIL_H} p(i) \log p(i) \quad (8)$$

The H_1^L in bit(s), related to the power in the bit(s) signal in H_1^L, can be used to determine the content of information symbol(s). Thus, the minimum amount of an image information in the overall domain of LIL(s) is $L \times H_1^L$, in addition, the information entropy-based data compression ratio in the domain of $1 - L$, ϑ_1^L, can be expressed as

$$\vartheta_1^L = \frac{LH_1^L}{n} \qquad (9)$$

Thus, the information entropy-based data compression ratio in the domain of $1 - L$ can be reduced to

$$\vartheta_{LIL_L}^{LIL_H} = \frac{(1+LIL_H-LIL_L)H_{LIL_L}^{LIL_H}}{n} \qquad (10)$$

In the FCMA of the corresponding histogram-based image (except the infinite $I(i)$ at LIL = 11), the H_1^L in (7) equals 0.8514 bit and the $H_{LIL_L}^{LIL_H}$ in (8) equals 0.1044 bit. Thus, the ϑ_1^L in (9) equals 0.8514 bit × 88 LILs/16384 pixels = 0.0046 bit/pixel and the $\vartheta_{LIL_L}^{LIL_H}$ in (10) significantly equals 0.1044 bit × 23 LILs/16384 pixels = 1.4656 × 10^{-4} bit/pixel in the image keyframe. In the FCMA of the HOS image (except the infinite $I(i)$ at LIL = 2 and the infinite $I(i)$ at LIL = 11), the H_1^L equals 0.3630 bit and the $H_{LIL_L}^{LIL_H}$ equals 0.0453 bit. Thus, the ϑ_1^L equals 0.3630 bit × 87 LILs/16384 pixels = 0.0019 bit/pixel and the $\vartheta_{LIL_L}^{LIL_H}$ significantly equals 0.0453 bit × 22 LILs/16384 pixels = 6.0828 × 10^{-5} bit/pixel in the image keyframe.

6. Conclusions and future work

Applying the FCMCA and the constraint $I(i)$ to the image classification, the corresponding histogram-based image information and the HOS image information have been investigated and the positions of important tissues in the specific MRI can be revealed. By the inverse self-information function, the LIL_L and LIL_H can be obtained. Thus, the lower ϑ_1^L and $\vartheta_{LIL_L}^{LIL_H}$ conformed to the H_1^L and $\vartheta_{LIL_L}^{LIL_H}$, respectively, in the information transmission can be performed. In future work, the more details concerning $\vartheta_{LIL_L}^{LIL_H}$ in the SPPS image will be investigated.

7. Acknowledgements

The authors would like to thank the research center of Dhurakij Pundit University for the research support.

Bibliography

Anderson, J. B., & Johannesson, R. (2005). *Understanding information transmission.* Wiley Inter-Science.

Homnan, B., & Benjapolakul, W. (2013). Comparison of the Spatial Resolution with the Histogram

Equalization-Obligated Linear Function on Structural Pixel Position Scheme. *Proceedings of ISCIT*. IEEE Xplore.

Jang, J. S. R., & Sun, C. T. (1997). *Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence.* Prentice Hall.

Reed, T. R. (2004). *Digital image sequence processing, compression and analysis*. CRC press.

Stallings, W. (2007*). Data and computer communications*. Pearson Prentice Hall.

Strannneby, D., & Walke, W. (2004). *Digital signal processing and applications*. Newnes.

Vasuda, P., & Satheesh, S. (2010). *Improved Fuzzy C-Means Algorithm for MR Brain Image*

Segmentation. International Journal on Computer Science and Engineering, 2(5), 1713–1715.