Computers and Mathematics with Applications 60 (2010) 2426–2432

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

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

A novel automatic microcalcification detection technique using Tsallis entropy & a type II fuzzy index

Mohanalin *, Beenamol, Prem Kumar Kalra, Nirmal Kumar

Department of Electrical Engineering, IIT Kanpur, UP-208016, India

ARTICLE INFO

Article history:
Received 18 August 2009
Received in revised form 12 August 2010
Accepted 12 August 2010

Keywords:
Tsallis entropy
Type II fuzzy index
Shannon entropy
Mammograms
Microcalcification

ABSTRACT

This article investigates a novel automatic microcalcification detection method using a type II fuzzy index. The thresholding is performed using the Tsallis entropy characterized by another parameter '*q*', which depends on the non-extensiveness of a mammogram. In previous studies, '*q*' was calculated using the histogram distribution, which can lead to erroneous results when pectoral muscles are included. In this study, we have used a type II fuzzy index to find the optimal value of '*q*'. The proposed approach has been tested on several mammograms. The results suggest that the proposed Tsallis entropy approach outperforms the two-dimensional non-fuzzy approach and the conventional Shannon entropy partition approach. Moreover, our thresholding technique is completely automatic, unlike the methods of previous related works. Without Tsallis entropy enhancement, detection of microcalcifications is meager: 80.21% Tps (true positives) with 8.1 Fps (false positives), whereas upon introduction of the Tsallis entropy, the results surge to 96.55% Tps with 0.4 Fps.

1. Introduction

Several articleshave been published highlighting the challenges existing in microcalcification cluster (Mcs) detection for early breast cancer diagnosis. Lately, Tsallis entropy (TE) based works have created a lot of interest [1]. It is proven that TE gives better thresholding results [2]. [3] compared the performance of traditionally used SE with TE and concluded that TE is far superior in detecting Mcs in mammograms. But there a histogram distribution technique was employed for calculating '*q*', which can lead to erroneous results when pectoral muscles are included. In the present study, a new technique based on type II fuzzy theory is proposed for calculating '*q*' optimally. The proposed approach has been tested on various images, and the results have demonstrated that the proposed TE approach outperforms the two-dimensional non-fuzzy approach and conventional SE partition approach.

The paper is organized as follows. Section 2 deals with the basics of TE and the proposed algorithm. Section 3 demonstrates the role of the type II fuzzy index in '*q*' identification. Sections 4 and 5 discuss the implementation of enhancement and the detection of Mcs respectively. The validation procedure adopted and the conclusions from our experiment are presented in Section 6.

2. The proposed algorithm

This section proposes TE based detection of Mcs using a type II fuzzy set. The proposed algorithm is shown in Fig. 1. The mammogram can be looked at as three individual objects fused together, namely, the picture background, the tissue background (including the fatty area) and Mcs (ROI). Generally in medical images the background remains black. Hence,

* Corresponding author.
E-mail address: mohanlee@iitk.ac.in (Mohanalin).

0898-1221/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2010.08.038

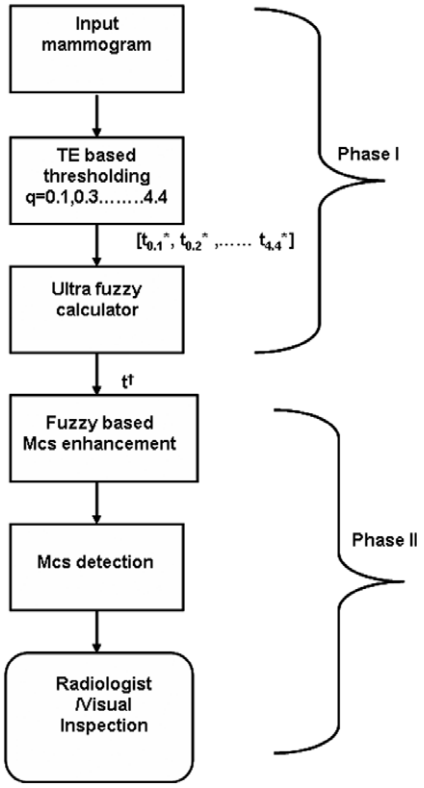

Fig. 1. Proposed algorithm using TE for detection of Mcs.

the intensity level of the background must be less than the average value of the intensity of the image. This object is of no significance and can be filtered off by employing the mean. The mean value '*k*' is calculated using Eq. (1):

$$k = \frac{1}{D} * \sum^{X} \sum_{m,n \in G}^{Y} g_{m,n} \tag{1}$$

with the following key:

M, N — the dimensions of the image;
G — the intensities larger than 100 for normal images and 10 for denser images;
D — the number of pixels;
g_{mn} — the grey level at coordinates m and n.

The Mcs will not be affected during this process, as generally the intensity of the Mcs is higher than the average of the image.

The probability distribution of grey levels in the mammogram can be written as $p_1, p_{k+1}, \ldots, p_N$. The tissue background (A) of the mammogram can be formulated as

$$\frac{p_k}{p_t - P_{k-1}}, \frac{p_{k+1}}{P_t - P_{k-1}}, \ldots, \frac{P_t}{P_t - P_{k-1}}.$$

The region of interest, i.e. the Mcs, in the mammogram (B) can be framed in terms of equations as follows:

$$\frac{p_{t+1}}{1 - P_t}, \frac{p_{t+2}}{1 - P_t}, \ldots, \frac{p_N}{1 - P_t}$$

$$P_{k-1} = \sum_{i=1}^{k-1} p_i$$

$$P_t = \sum_{j=k}^{t} p_j.$$

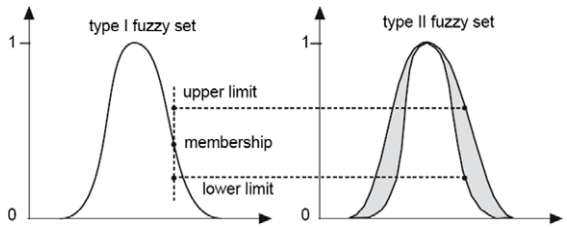

Fig. 2. Type II fuzzy set.

The entropies for the tissue background (A) can be written as follows:

$$S_q(A) = \left(\frac{1}{q-1}\right) * \left(1 - \left(\sum_{i=k}^{t} \frac{p_i}{P_t - P_{k-1}}\right)^q\right). \tag{2}$$

Similarly for the ROI, the entropy measure can be written as

$$S_q(B) = \left(\frac{1}{q-1}\right) * \left(1 - \left(\sum_{i=t+1}^{N} \frac{p_i}{1 - P_t}\right)^q\right). \tag{3}$$

The maximum separation between A and B can be found using a traditional energy maximization technique, which gives the optimal value of 't^*':

$$t^* = \text{argmax}\{S_q(A) + S_q(B)\}. \tag{4}$$

Since TE possess the nonadditive property,

$$S_q(A + B) = S_q(A) + S_q(B) + (1 - q)(S_q(A) * S_q(B)). \tag{5}$$

Eq. (4) is the optimal threshold which separates Mcs from the tissues ideally. But the information contained in Eq. (4) is fuzzy. This is due to the nature of 'q'. It can be any real positive value. To avoid the fuzziness we are introducing a type II fuzzy index where the ultra-fuzziness is calculated at each 'q'.

3. The type II fuzzy index

Hamid R. Tizhoosh [4] has used a type II fuzzy set for thresholding. He has explained that a type I fuzzy item is still fuzzy, and termed this ultra-fuzziness. i.e., the membership function is crisp. A type II fuzzy set is shown in Fig. 2. This phase is intended to fuzzify the fuzzified image and also to find the fuzzy number which indicates to what extent the fuzzified image is fuzzy.

The type II fuzzy index is defined as

$$\gamma(A) = \frac{1}{M * N} \sum_{g=0}^{L-1} h(g) * [\mu_U(g) - \mu_L(g)] \tag{6}$$

$$\mu_U(g) = \left[\mu_A(g)^{1/\alpha}\right]$$

$$\mu_L(g) = \left[\mu_A(g)^{\alpha}\right]$$

with $\alpha \epsilon (0, 2]$, with the following key:

$\mu_U(g)$ – the upper membership function;
$\mu_L(g)$ – the lower membership function;
$\mu_A(g)$ – the Gaussian membership function;
$h(g)$ – the histogram of the image;
$M * N$ – the dimensions of the image.

Mendel has introduced an index of fuzziness to measure the vagueness of a fuzzy set. The complete details of the type II fuzzy set can be seen in the article written by him [5]. Eq. (5) calculates the index of vagueness of the image under study for various 'q' values. The optimal threshold can be defined as follows:

$$t^* = \text{argmax}\{\gamma(A_i)\} \tag{7}$$

$i = i, \ldots, n$.

Since 'q' values are not precise, several preset values are defined. In this approach, the values of 'q' used are 0.1, 0.3, 0.5, 0.7, 0.9, 2, 3, 4.1, 4.3, 4.4. The values are not restricted and any value can be used. The above mentioned values are selected

from experience, indicating that other values will not play a crucial role. Otherwise there is no specific reason behind this. Instead, one can set out to find the ultra-fuzziness at each 'q' in some interval (e.g., [0, 5] etc.). [3] discusses the details of the procedure used to identify the 'q' values defined and the readers are advised to read [3] for complete details. Eq. (7) computes the ultra-fuzziness (fuzzy number) for the 10 different thresholds calculated using Eq. (4). The threshold which possesses maximum ultra-fuzziness is identified using the maximum operation. By this method, the optimal threshold t^* is selected, which corresponds to the optimal 'q'.

4. Enhancement of Mcs

Mammograms do not have well defined shapes and they are fuzzy in nature. The fuzzy entropy principle is optimally suited for enhancing mammograms. Fuzzification involves transformation of the values of the intensity to an interval between 0 and 1. This can be done by using any appropriate fuzzy membership function. The function defined is used to locate the intensities of Mcs. A Gaussian membership function is often used because of its simplicity and robustness. The selection of a crossover point can be viewed as an object–background classification problem. Hence thresholding techniques can be applied. Since Mcs have more intensity than the tissue [6], it is obvious that the fuzzy region of the function must be in the range from the mean intensity to the maximum intensity of the mammogram. In other words, the proposed algorithm gives maximum membership value to intensities of Mcs higher than the threshold t^* and suppresses the remainder. The Gaussian function can be written as

$$\pi_{m,n} = \exp\left(-\frac{|g_{mn} - \max N|^2}{f_h^2}\right) \tag{8}$$

with the following key:

$\max N$ — the maximum intensity value;
f_h^2 — the bandwidth of the Gaussian membership function.

The bandwidth of the Gaussian function can be calculated using the following equation:

$$f_h = \max\left\{(t^* - k), (\max N - t^*)\right\}$$

where t^* is the threshold found using (7).

In this way, it suppresses the region which does not belong to the ROI and enhances the remaining region. Further, local geometrical information is utilized for computing the non-uniformity of the image. i.e., local variances are employed for this, and can be calculated as follows:

$$\mu_{m,n} = \frac{1}{W_x * W_x} \sum_{j=1}^{W_x^2} g_{m,n} \tag{9}$$

$$\sigma_l^2 = \frac{1}{W_x * W_x} \sum_{j=1}^{W_x^2} [g_{m,n} - \mu_{m,n}] \tag{10}$$

with the following key:

$W_x * W_x$ — the window;
$\mu_{m,n}$ — the local mean;
σ_l^2 — the local variance.

The non-uniformity factor calculated will be normalized by using the optimum threshold found:

$$V_l = \begin{Bmatrix} \frac{\sigma_l^2}{t^*} & \sigma_l^2 \le t^* \\ 1 & \text{otherwise} \end{Bmatrix}. \tag{11}$$

The enhanced mammogram can be obtained by using the following formula:

$$G_{\text{enh}} = \pi_{m,n} * V_l * \max N \tag{12}$$

with the following key:

$\pi_{m,n}$ — the fuzzified mammogram;
G_{enh} — the enhanced mammogram.

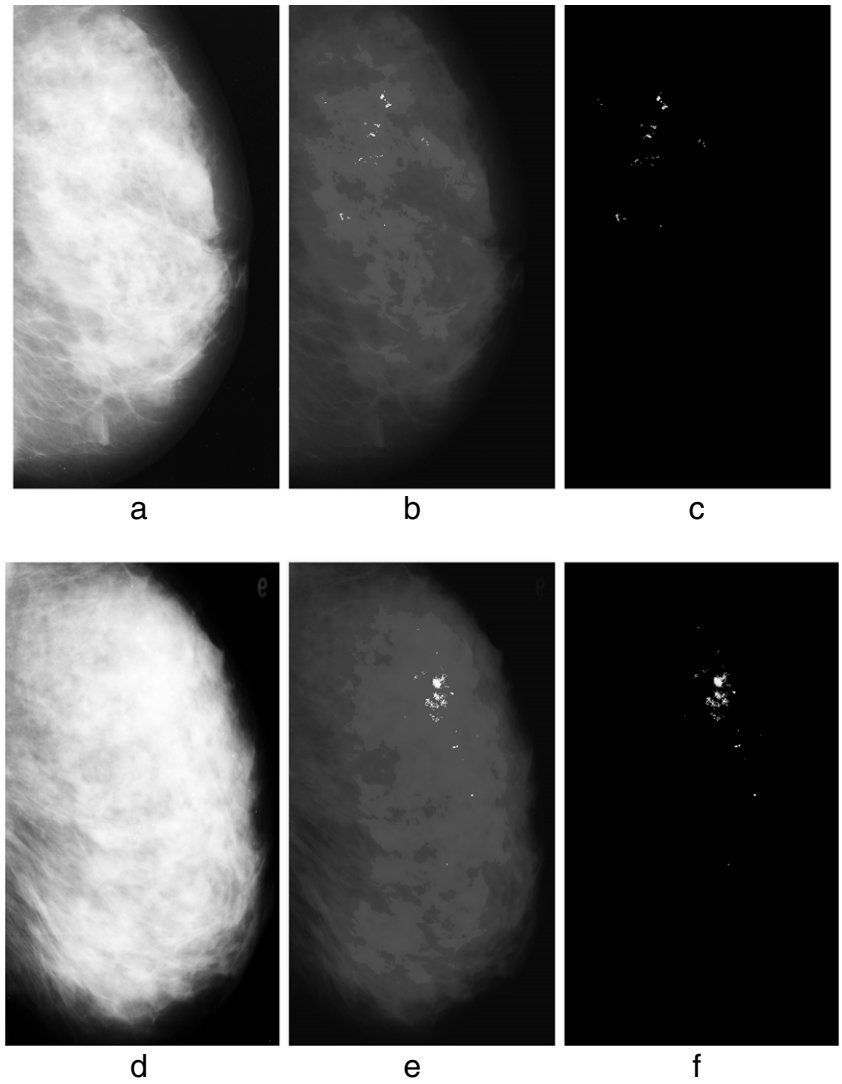

Fig. 3. Detection results from the experiment: (a) mdb236, (b) output with the Mcs enhanced, (c) output with the Mcs extracted; (d) mdb216, (e) output with the Mcs enhanced, (f) output with the Mcs extracted.

5. Detection of Mcs

A simple iterative averaging technique is employed to detect the Mcs. This technique selects the threshold iteratively. First, the input image is averaged to find an initial threshold T_0. This threshold is then used to detect the Mcs, and the newly derived image is represented as a new input image. The same procedure can then be repeated until the change of the threshold values ($abs(T - T_0)$) is less than a preset value ξ. In this experiment, ξ is set as 5. Results are shown for randomly picked mammograms in Fig. 3. Fig. 3 shows the enhanced images and the Mcs detected from them.

5.1. The important steps of the proposed approach

1. Use the TE to find the threshold t^* for all defined 'q' values.
2. Calculate the image histogram.
3. Initialize the position of the membership function.
4. Fuzzify the image using the Gaussian membership function with t^*.
5. Calculate in each pixel position the amount of ultra-fuzziness using (6).
6. Determine the t^* corresponding to q_{opt} with maximum ultra-fuzziness T.
7. Fuzzify the image using t^*, to enhance the Mcs.
8. Use the iterative thresholding technique to extract the Mcs alone.

Table 1
Tps and Fps calculated for MIAS and UCSF individually and combined.

Name of database	No. of images	No. of Mcs	Tps detected	Tp rate in (%)	No. of Fps	Fps per image
MIAS	50	32	30	93.75	26	0.52
UCSF	197	142	138	97.18	73	0.37
Combined analysis	247	174	168	96.55	99	0.4

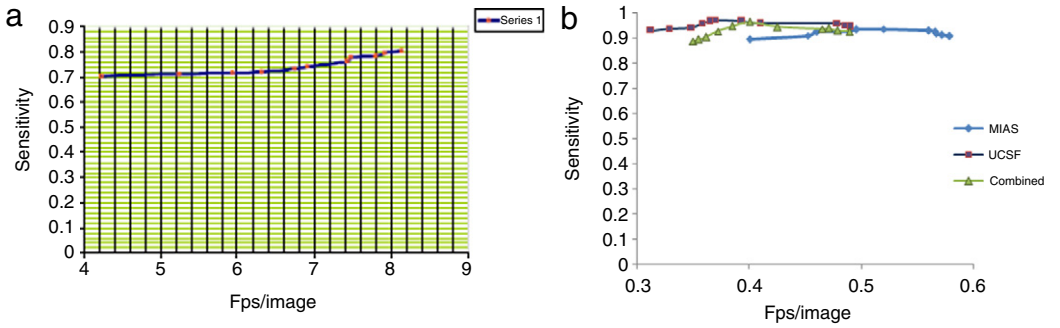

Fig. 4. FROC curves: (a) before using the TE; (b) after inclusion of the proposed algorithm.

Table 2
Comparing the performance of the proposed type II fuzzy based algorithm with other algorithms.

References						
Features	Exp no: 2	[7]	[8]	[9]	[10]	[11]
Breast region extraction	Manual	Genetic algorithm	N/A	Segment	Iterative thresholding	N/A
Tsallis parameter	Yes	Nil	Nil	Nil	Nil	Nil
ROI detection	Annotated by radiologists	Asymmetry approach	Morphological descriptors	Annotated by radiologists	Discrete wavelet transform	Annotated by radiologists
Enhancement	Fuzzy based	Gradient based	Contrast	Wavelet	Intensity remapping method	N/A
Mcs detection	TE	Median filter	Hybrid intelligent system	Neural network	Filling dilation	Seed growing
Class	Benign; malignant	Benign; malignant	Benign; malignant	Benign; malignant	Benign; malignant	Benign; malignant
Evaluation	FROC	ROC	ROC	FROC	FROC	Classification rate
Result	97.7% Tps & 0.26 Fps	Az = 0.906	Az = 0.81 & 0.80	88% Tps & 2.15 Fps	96.9% Tps 0.2 Fps	93.55%

6. FROC analysis for validating the detection procedure

The FROC analysis was conducted for 247 images with and without Mcs, selected from UCSF and MIAS. Out of 174 Mcs, 168 Mcs were detected correctly, While 6 true Mcs were missed, 99 false Mcs were mistaken for true ones. So the Tp rate is 96.55% (168/174), with the number of the Fps per image 0.4 (99/247). Further, the proposed algorithm was able to detect the Mcs with a high precision, with a low Fp rate. The FROC curve shown in Fig. 4 confirms this. The Tp rate obtained is 96.55% with only 0.4 Fps per image. [6] gives 97% Tps but 3 Fps/image and without the TE it led to 80.21% Tps with 8.1 Fps, which shows our algorithm to be much better. Table 1 displays the Tps and Fps calculated for MIAS and UCSF individually and collectively. Table 2 compares the performance of the proposed approach with those of other popular algorithms. From Table 2, we can conclude that the proposed algorithm does indeed have the potential to identify Mcs with far superior results.

References

[1] J. Mohanalin, P.K. Kalra, Nirmal Kumar, Tsallis entropy based microcalcification segmentation, ICGST-GVIP Journal 9 (I) (2009).
[2] M. Portes de Albuquerque, et al., Image thresholding using Tsallis entropy, Pattern Recognition Letters 25 (9) (2004) 1059–1065.
[3] J. Mohanalin, Prem Kumar Kalra, Nirmal Kumar, An automatic method to enhance microcalcifications using normalized Tsallis entropy, Elsevier Signal Processing 90 (3) (2010) 952–958.
[4] Hamid R. Tizhoosh, Image thresholding using type II fuzzy sets, Pattern Recognition 38 (12) (2005) 2363–2372.
[5] J.M. Mendel, R.I. John, Type-2 fuzzy sets made simple, IEEE Transactions on Fuzzy Systems 10 (2002) 117–127.
[6] Heng-Da Cheng, Yui Man Lui, Rita I. Freimanis, A novel approach to microcalcification detection using fuzzy logic, IEEE Transactions on Medical Imaging 17 (3) (1998) 442–450.

[7] K. Thangavel, M. Karnan, Computer aided diagnosis in digital mammograms: detection of microcalcifications by meta-heuristic algorithms, ICGST-GVIP Journal 5 (7) (2005).
[8] A. Papadopoulos, D.I. Fotiadis, A. Likas, Characterization of clustered microcalcifications in digitized mammograms using neural networks and support vector machine, Artificial Intelligence in Medicine 34 (2005) 141–150.
[9] P. Delogu, M.E. Fantacci, A.P. Martinez, A. Retico, A. Stefanini, A. Tata, A scalable system for microcalcification cluster automated detection in a distributed mammographic database, in: IEEE Nuclear Science Symposium Conference Record, 2005, pp. 1530–1534.
[10] W. Xu, S. Xia, H. Duan, A Novel Computer Aided Diagnosis System of the Mammogram, Springer-Verlag, 2006, pp. 639–644.
[11] N.A. Mat Isa, S. Sabaruddin, U.K. Ngah, K.Z. Zamli, M. Mohd Noor, The potential use of modified seed-based region growing technique for automatic detection of breast microcalcifications and tumour areas, Jurnal Teknologi 44 (2006) 151–164.