Can entropy-based image alignment metrics offer improved image aggregation of tissue density for mammographic risk assessment?

Final Report for CS39440 Major Project

Author: Laura Collins (lac32@aber.ac.uk)
Supervisor: Dr. Neil Mac Parthalin (ncm@aber.ac.uk)

4th March 2016
Version: 1.0 (Draft)

This report was submitted as partial fulfilment of a BSc degree in Artificial Intelligence and Robotics (inc Integrated Industrial and Professional Training) (GH7P)

Department of Computer Science
Aberystwyth University
Aberystwyth
Ceredigion
SY23 3DB
Wales, UK

Declaration of originality

In signing below, I confirm that:

- This submission is my own work, except where clearly indicated.
- I understand that there are severe penalties for Unacceptable Academic Practice, which can lead to loss of marks or even the withholding of a degree.
- I have read the regulations on Unacceptable Academic Practice from the University's Academic Quality and Records Office (AQRO) and the relevant sections of the current Student Handbook of the Department of Computer Science.
- In submitting this work I understand and agree to abide by the University's regulations governing these issues.

Name ....

Date ....

Consent to share this work

In signing below, I hereby agree to this dissertation being made available to other students and academic staff of the Aberystwyth Computer Science Department.

Name ....

Date ....

Acknowledgements

I would like to thank my Supervisor Neil for his constant help and guidance throughout this project.

Ryan for being my constant sound-board throughout the process, always happy to lend an ear when I needed to work through an issue, or bounce programming ideas off of someone.

Harry, Fangyi

Charlie

Abstract

Include an abstract for your project. This should be no more than 300 words.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Background & Objectives

The Project is concerned with the alignment of multiple images using an image-alignment technique called Congealing [5]. The paper describes a method which utilises the Congealing algorithm to align both MNIST handwriting data and MRI Scans by reducing the pixel-wise uncertainty across a collection of images. This project aims to go further, aligning mammography scans and reducing the pixel-wise uncertainty using a number of different fuzzy entropy methods.

1.1 Background

In preparation for my Major Project extensive research was undertaken into the Fuzzy Entropy alignment metrics that would be chosen and implemented in the Congealing algorithm to align mammography scans. S. Al-sharhan's paper 'Fuzzy Entropy: a Brief Survey' [1] was instrumental in a brief comparison between Fuzzy Entropy metrics - allowing a simple way to compare and contrast mathematical differences and implementations.

Researchers have implemented many variations of the Congealing algorithm, with success in varying areas - however very little to no work has been done implementing the algorithm to assess mammography scans. Cox's 'Least Squares Congealing' [2] was quickly disregarded given the project preference to select entropy-based alignment techniques, however did alert to the issue of performance for the project.

The motivation behind choosing mammograms as the input data of choice was an interest into how computer systems and machine learning can help in the medical sector.

1.2 Analysis

1.2.1 Task composition

1.2.1.1 Pixel Membership

From the analysis of the planned Fuzzy Entropy algorithms, one major task to be undertaken would be to calculate the membership of each pixel. Membership stems from Fuzzy set theory **Check this?** which acknowledges that an element can be part of more than one group to a certain degree, not just one or the other. One common example of this is listing someone as 'Short', 'Average' or 'Tall' in height. If a tall person is someone over 6 feet in height, would a person who measured 5foot 11inches not be classified as tall? In fuzzy set theory, they would be be a certain degree of tall, and a certain degree of average, with the highest membership likely to win out when categorising their height.

There are two common methods to modeling degrees of membership. The first is to manually define the categorty boundaries, so in the case of trapezium functions, the two bases and the two shoulders as in Figure x **include a diagram here showing?**. The other solution would be to iterate over the values you have and to computationally build the an even distribution throughout your membership functions. Whilst this is the preferred method for being dynamic in it's calculations, it is also more computationally expensive as pre-processing of the image would have to be completed before the Congealing algorithm could be run.

Taking this into account, for grey-level pixel values, ranging from 0 (black) to 255 (white), three trapezium functions would be sufficient, therefore modeling 'Low', 'Medium' and 'High' grey-level values. The bases and shoulders would be statically defined by the User, however would be open for editing between running the Congealing code for experimentation purposes. For Non-Probabilistic entropy the highest membership for each pixel from each of the three trapeziums would be taken as the membership degree.

1.2.2 Research questions

- Does the use of Fuzzy Entropy alignment metrics improve the alignment of mammograms?
- Do clinicians / radiographers / mammographers find the output at all useful?
- What advantages / disadvantages does each fuzzy entropy alignment metric entail?

1.3 Research Method

Talk about lit review here

For this project, a literature review was undertaken to assess the work completed by researchers in the fields of Fuzzy Entropy and image alignment methods.

1.3.1 Entropy

1.3.2 Fuzzy Entropy

There are three different types of uncertainty:

- Probabilistic
- Possibilistic
- Indiscernibility

1.3.3 Image Alignment methods

Some work has already been undertaken to investigate image alignment using Fuzzy Entropy metrics, however typically they are computationally costly, and therefore slow to run. This project will be invesitgating whether there are simpler, more light-weight fuzzy entropy metrics which could be implemented, for more everyday use in image alignment. It will also be investigated if, and further how, the outputs of these alignments differ per each fuzzy entropy metric.

1.3.3.1 Congealing

Chapter 2

Experiment Methods

2.1 Overview

2.2 Implementation tools

MATLAB

2.3 Algorithms

2.3.1 Shannon Entropy

2.3.2 Non-Probabilistic Entropy

2.3.2.1 Fuzzy entropy description

De-Luca & Termini fuzzy entropy algorithm [3] is considered to be the first to build upon Shannon entropy. Their implementation takes into account a set of data, along with their various membership degrees.

$$H_A = -K \sum_{i=1}^{n} \{\mu_i \log(\mu_i) + (1 - \mu_i) \log(1 - \mu_i)\} \tag{1}$$

Al-sharhan et al's paper compiling several Fuzzy Entropy algorithms [1] contains a methodical,in-depth derivation of their algorithm, and has been instrumental in building my knowledge on the algorithm in question.

We will assume $-K$, the positive constant, is defined as $\frac{1}{n}$ as outlined in [3].

2.3.2.2 MATLAB implementation

After some research into current implementations of Fuzzy Entropy algorithms in MATLAB, it was concluded the best approach would be to implement De-Luca & Termini's algorithm from scratch. This entailed creating a membership class, which computes the grey-level membership of each pixel in the mean image (calculated from a set of input images).

This array of pixel memberships is fed into a 'De Luca' function where it is iteratively passed into latter part of equation 1 (after $\sum$). The output array is then summed and multiplied by $\frac{1}{n}$ as defined in Section 2.3.2.1. The final mean pixel entropy is calculated by taking the image entropy and dividing by the number of pixels in the image.

This is all relatively straight forward to implement in MATLAB, as it is designed to run mathematical equations.

2.3.2.3 Technical challenges

The main technical challenge for this implementation is ensuring maximum optimisation to keep running times to a minimum. Leveraging MATLAB's own functions for the membership saves a lot of time and lines of code, however it's been important to check what they call from within. One membership function was redrawing the trapeziums every time it was called, significantly slowing down the process - reducing the amount of times the initial function was called helped reduced the run-time by over 60seconds.

Another technical challenge faced whilst implementing the De Luca & Termini algorithm, isn't directly tied to the implementation of their specific equation, but more of my lack of experience in MATLAB, slowing down the programming rate. It has indeed been a steep learning curve, getting to grips with standard error messages, the debugger tool and knowing which 'Toolboxes' are needed to run specific MATLAB functions.

Finally, as can be ascertained from Figure 2.2, when writing the 4 seperate scans into 1 larger file, somewhere the images get rotated. This will be a reoccurring issue through the 3 Fuzzy entropy implementations, however as this is the first I will note it here. I think this is caused thanks to the swapping of the height and width values, however upon initial inspection of the file writing function, it is not clear as to which line is causing this issue. This issue has been marked low priority in the short term, due to all the scans being rotated in this fashion, and as such all have the same orientation. This means the congealing algorithm can work with no issues upon these images, the rotation is more merely an aesthetic issue.

2.3.2.4 Results

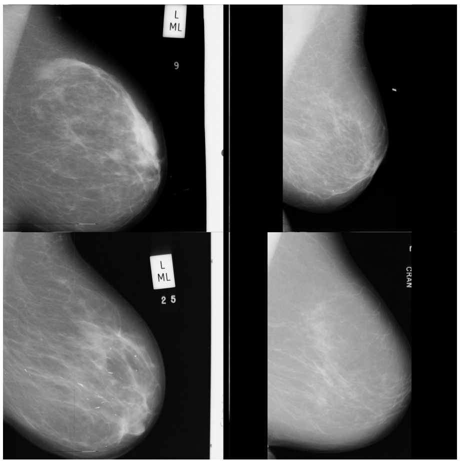
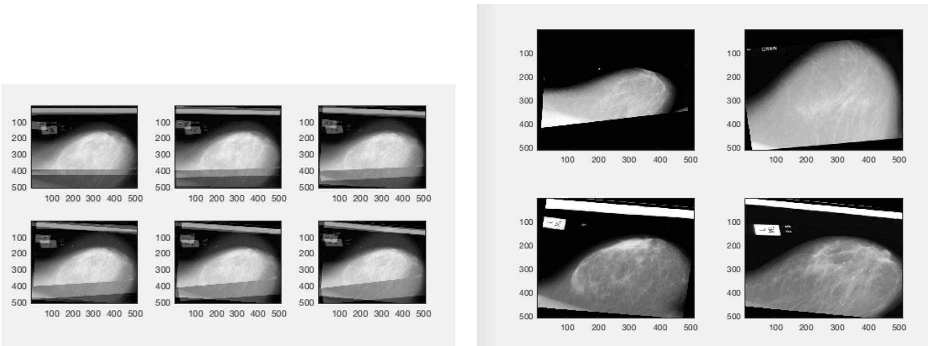


Figure 2.1: 4 input images of BI-RADS I classification



(a) Mean image over each iteration (b) Adjusted original input images

Figure 2.2: Output of 5 congealing iterations

2.3.3 Entropy results

Iteration	Entropy
1	0.050519
2	0.043925
3	0.035679
4	0.029035
5	0.026194

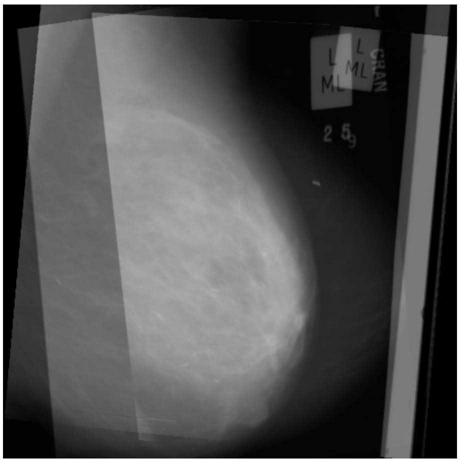


Figure 2.3: Final mean image after 5 iterations (bottom-right most in Figure 2.2)

2.3.4 Time to Run

Start Profiling Run this code: Profile time: 384 s

Profile Summary

Generated 16-Mar-2016 11:04:28 using performance time.

Function Name	Calls	Total Time	Self Time*	Total Time Plot (dark band = self time)
deLucaFuzzy	236	204.912 s	202.820 s	
membership	236	173.307 s	169.140 s	
incrTrans	20	5.286 s	0.475 s	
computeXfrmImg	231	3.966 s	1.252 s	
interp2	251	2.922 s	2.650 s	
imwrite	236	2.605 s	0.132 s	
imagesci/private/writepnm	236	2.353 s	0.102 s	
testCongeal	1	2.299 s	0.154 s	
imagesci/private/writepnm>writepgm	236	2.205 s	0.108 s	
imagesci/private/writepnm>write_raw_data	236	1.820 s	1.820 s	
binaryCongeal	1	1.617 s	0.062 s	
addpath	244	1.432 s	0.050 s	
path	244	1.363 s	0.867 s	

Figure 2.4: Snapshot of run-time statistics

2.3.5 Hybrid Entropy

2.3.5.1 Fuzzy entropy description

Pal and Pal introduced Hybrid Entropy [6] in 1992 to help combat the issues faced by Non-Probablisitc entropy - mainly that it does not model probabilistic uncertainty. The name 'Hybrid' stems from the fact it models both Probabilistic and Possibilistic (fuzziness) uncertainty.

This could probably be covered in Lit review.

2.3.5.2 MATLAB implementation

Due to reasons covered in the Subsubsection 2.3.5.3, Hybrid Entropy membership was implemented using 2 trapeziums covering 2 fuzzy sets.

Two arrays are then fed into the Hybrid Entropy function - listing all the pixel membership values from the low trapezium, and the high trapezium. The final entropy is taken as a comparison between the low and high fuzzy sets.

2.3.5.3 Technical challenges

Whilst Hybrid Entropy utilises a membership function, much like Non-Probabilistic entropy, it was derived to work with binary entropy, not the ternary membership modeled for Non-Probabilistic. Because of the binary nature, the equation uses 'inversion' to depict if not this fuzzy set, then must belong to the other.

Experimentation was done as to whether the equation could be adapted in such a way to continue using three separate membership trapeziums - low, medium and high grey-level values.

Initial ideas - check email between me and neil

Logic would dictate that if the comparison of two fuzzy sets works, then to compare the low fuzzy set to the medium, the medium to the high and the high to the medium should work.

For example:

In theory, calculating E_0 and E_1 for each trapezium, calculating the hybrid entropy for each, and then combining them, should work:

$$E_0 = \frac{1}{No.\_of\_pixels\_in\_low\_trapezium} \sum_{i=1}^{n} (1 - Low\mu_i) exp(Low\mu_i) \tag{2}$$

$$E_1 = \frac{1}{No.\_of\_pixels\_in\_low\_trapezium} \sum_{i=1}^{n} Low\mu_i exp(1 - Low\mu_i) \tag{3}$$

Where $Low\mu$ is the membership of the pixels in the low fuzzy set.

$$H_{hy} = -p_0 \log_{10}(1 - E_0) - p_1 \log_{10}(E_1) \tag{4}$$

Where

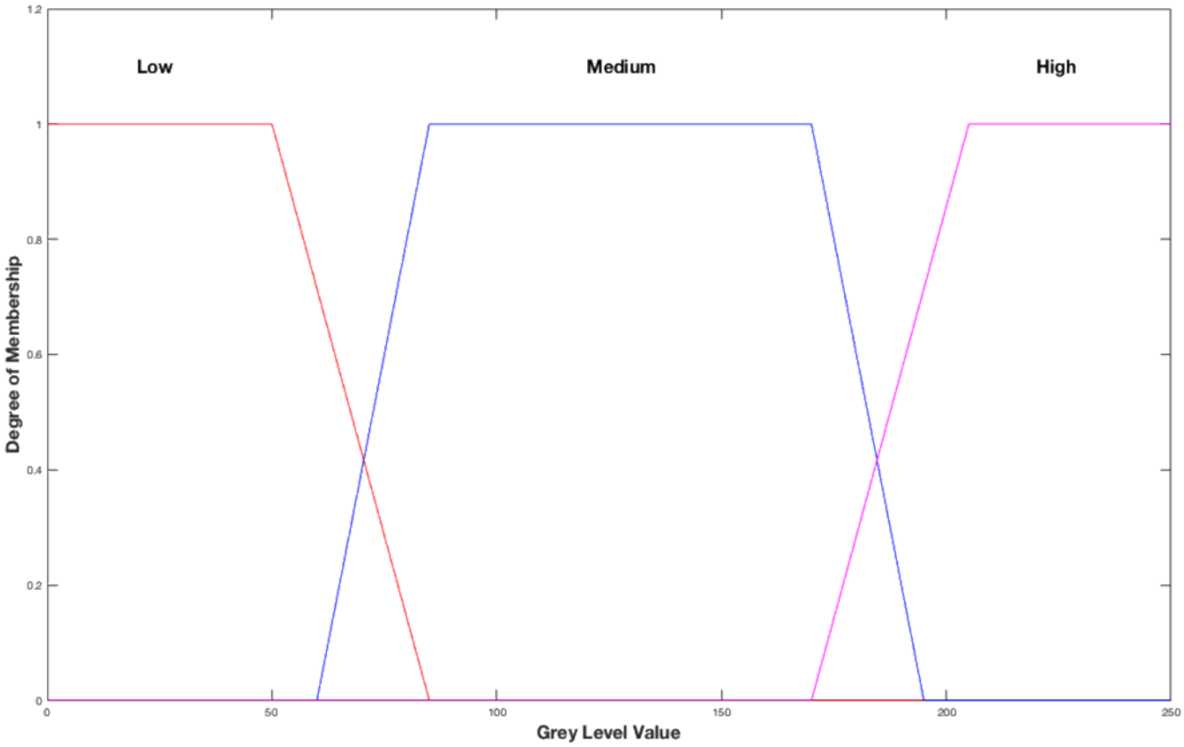


Figure 2.5: 3 fuzzy set trapeziums

$p_0 = \frac{No.\_of\_pixels\_in\_low\_trapezium}{No.\_of\_pixels\_in\_low\_trapezium+med\_trapezium}$

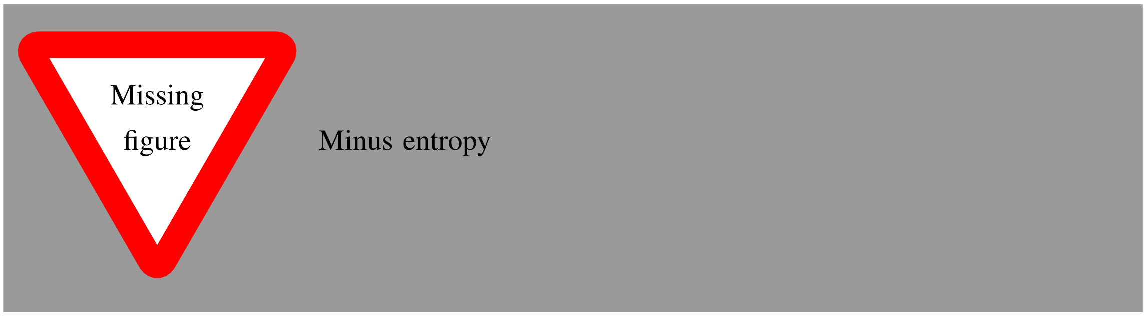
and

$p_1 = \frac{No.\_of\_pixels\_in\_med\_trapezium}{No.\_of\_pixels\_in\_low\_trapezium+med\_trapezium}$

This was done for all 3 trapeziums, then combined and divided by 3 (for the mean entropy). As the result for each trapezium should be between 0 and 1 (as each is an entropy value), then combining them should be no issue. However this was not the case.

First of all, the hybrid equation output was deemed to be ‘NaN’ - something which generally occurs when attempting to divide by 0. Anomalous outputs from the high trapezium was to be expected, as there are very few pixels which fall within the range nearer the white end of the grey-level scale. This was mitigated by setting the output equal to 0, in effect ignoring any output from the highest fuzzy set.

After this mitigation, the third and fourth iteration had suitable entropy values, however the fifth entropy value was a negative - something not possible in terms of entropy.



It was concluded that the implementation of three fuzzy sets within Hybrid Entropy would not be realistic within the remaining timeframe of the project, and the membership for Hybrid Entropy was redefined to the concept of 2 fuzzy sets, as set out by Pal and Pal.

2.3.5.4 Results

2.4 Software

2.4.1 Methodology

An adapted Scrum methodology has been undertaken for this project. This has been supported by the tool available at `taiga.io` - a beta web app.

- Burn down chart
- User stories
- Retrospectives
- Daily standup

2.4.2 Design

- CRC cards

2.4.3 Implementation

2.4.4 Testing

Chapter 3

Results and Conclusions

Chapter 4

Critical Evaluation

Appendices

Appendix A

Third-Party Code and Libraries

1.1 Congealing Code

The project focused on extending the existing Congealing Code implemented by Learned Miller et al in 2005. A Congealing demo is available on the Congealing website [4] which is open for experimentation. The original demo code was modified and extended to be able to read in mammograms and to work with 2 Fuzzy Entropy algorithms.

Appendix B

Ethics Submission

2.1 Ethics Application Number: 3958

AU Status
Undergraduate or PG Taught

Your aber.ac.uk email address
lac32@aber.ac.uk

Full Name
Laura Collins

Please enter the name of the person responsible for reviewing your assessment.
Reyer Zwiggelaar

Please enter the aber.ac.uk email address of the person responsible for reviewing your application
rrz@aber.ac.uk

Supervisor or Institute Director of Research Department
cs

Module code (Only enter if you have been asked to do so)
CS39440

Proposed Study Title
Entropy based metrics for joint image alignment

Proposed Start Date
25th January 2016

Proposed Completion Date
4th May 2015

Are you conducting a quantitative or qualitative research project?
Mixed Methods

Does your research require external ethical approval under the Health Research Authority?
No

Does your research involve animals?
No

Does your research involve human participants?
Yes

Are you completing this form for your own research?
Yes

Does your research involve human participants?
Yes

Institute
IMPACS

Please provide a brief summary of your project (150 word max)
I will be investigating the use of Congealing multiple MIAS dataset mammograms using several fuzzy entropy alignment metrics. If time permits I plan on speaking to a specialist (radiologist) to determine whether the output mean images of the congealing process are of any significant use to the research into breast cancer detection.

I can confirm that the study does not involve vulnerable participants including participants under the age of 18, those with learning/communication or associated difficulties or those that are otherwise unable to provide informed consent?
Yes

I can confirm that the participants will not be asked to take part in the study without their consent or knowledge at the time and participants will be fully informed of the purpose of the research (including what data will be gathered and how it shall be used during and after the study). Participants will also be given time to consider whether they wish to take part in the study and be given the right to withdraw at any given time.
Yes

I can confirm that there is no risk that the nature of the research topic might lead to disclosures from the participant concerning their own involvement in illegal activities or other activities that represent a risk to themselves or others (e.g. sexual activity, drug use or professional misconduct). Should a disclosure be made, you should be aware of your responsibilities and boundaries as a researcher and be aware of whom to contact should the need arise (i.e. your supervisor).
Yes

I can confirm that the study will not induce stress, anxiety, lead to humiliation or cause harm or any other negative consequences beyond the risks encountered in the participant's day-to-day lives.
Yes

Please include any further relevant information for this section here:

Where appropriate, do you have consent for the publication, reproduction or use of any unpublished material?
Yes

Will appropriate measures be put in place for the secure and confidential storage of data?
Yes

Does the research pose more than minimal and predictable risk to the researcher?
No

Will you be travelling, as a foreign national, in to any areas that the UK Foreign and Commonwealth Office advise against travel to?
No

Please include any further relevant information for this section here:

If you are to be working alone with vulnerable people or children, you may need a DBS (CRB) check. Tick to confirm that you will ensure you comply with this requirement should you identify that you require one.
Yes

Declaration: Please tick to confirm that you have completed this form to the best of your knowledge and that you will inform your department should the proposal significantly change.
Yes

Please include any further relevant information for this section here:

Appendix C

Code Examples

Annotated Bibliography

[1] S. Al-Sharhan, F. Karray, W. Gueaieb, and O. Basir, "Fuzzy entropy: a brief survey," in *Fuzzy Systems, 2001. The 10th IEEE International Conference on*, vol. 3. IEEE, 2001, pp. 1135–1139. [Online]. Available: http://dx.doi.org/10.1109/fuzz.2001.1008855

Paper outlining the different implementations of Fuzzy Entropy, of which 3 will be selected and focused on during this Project.

[2] M. Cox, S. Sridharan, S. Lucey, and J. Cohn, *Least squares congealing for unsupervised alignment of images*, Jun 2008, p. 18.

A disregarded adaption of the Congealing algorithm - however was useful in highlighting performance issues in the original algorithm. Something which was near continuously faced when implementing heavier fuzzy entropy alignment metrics.

[3] A. De Luca and S. Termini, "A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory," *Information and Control*, vol. 20, no. 4, p. 301312, May 1972.

Some annotation here

[4] E. Learned-Miller, "The congealing page." [Online]. Available: https://people.cs.umass.edu/~elm/congealing/

[5] E. G. Learned-Miller, "Data driven image models through continuous joint alignment," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 2, pp. 236–250, Feb. 2006. [Online]. Available: http://dx.doi.org/10.1109/tpami.2006.34

Learned-Miller's original Congealing method is the basis for this Project - however I am looking to further extend the alignment capabilities using fuzzy entropy metrics, rather than standard Shannon entropy as currently implemented. This paper was extremely useful for understanding of the basic concepts behind it, and will be a good reference guide throughout the project.

[6] N. R. Pal and S. K. Pal, "Higher order fuzzy entropy and hybrid entropy of a set," *Information Sciences*, vol. 61, no. 3, p. 211231, Jun 1992.