

# **Adaptive binarization of legacy ionization chamber cosmic ray recordings**

**Dissertation submitted in partial fulfilment of the requirements  
for the degree Master of Science in Computer Science at the  
Potchefstroom Campus of the North-West University**

Andre Steyn  
20535341

Supervisor:  
G.R. Drevin

October 2012

# Acknowledgments

I would never have been able to finish this dissertation without the support and guidance of some truly remarkable people.

Firstly, I would like to thank my parents, Maartin and Igna Steyn, for providing me with the opportunity to obtain a masters degree and for supporting me financially throughout my years of study.

I would like to thank my supervisor, Prof. Gunther Drevin, for allowing me the freedom to follow my own methods while ensuring that I do not stray from my final objective and also for being an abundant source of advice.

I would like to thank Prof. Magda Huisman, who has taught me all I know about how to conduct research and how to cope with the pressure of writing a dissertation.

I would also like to thank Kobie Fourie, who always has a willing shoulder and is invaluable when administrative challenges present themselves.

Finally, I would like to thank my girlfriend, Anli van den Berg, for all her support and motivation. She was always willing to listen to problems and ideas and played a big role in the initial language editing of this dissertation.

# Abstract

In the 1930s, the Carnegie Institute in Washington DC initiated the construction of cosmic ray observation centres around the world. Cosmic ray activity was recorded using the model C cosmic ray ionization chamber which uses a Lindemann electrometer. Seven of these chambers were constructed at seven stations around the world.

These chambers recorded cosmic ray data by projecting the shadow of the electrometer needle onto a continuously moving strip of 60 mm photographic paper. Hour markers were recorded by dimming the lamp for three minutes at the start of each hour, while also grounding the ionization chamber. By grounding the ionization chamber the electrometer needle was returned to the zero position. The photographic paper moved about 25 mm an hour. Approximately 114 station-years of data was recorded between 1935 and 1960 (Hardy, 2006).

It is important to digitize these recordings in order to preserve the data for further study of cosmic rays from this time period. This digitization process consists of binarizing digital images of the photographic strip to extract the cosmic ray data. By binarizing these images the data is recorded in an easily usable format for future research.

This study focuses on extraction of the cosmic ray data using an adaptive binarization method that is able to cope with a wide variety of images, ranging from images that are almost too bright to distinguish the data lines from the background, to images that are too dark to distinguish the data lines at all.

This study starts off with a brief explanation of cosmic rays, how these were recorded before the 1950s and how the rays are recorded today.

**Two research methodologies were used to create a method to adaptively binarize and extract data from the historic cosmic ray recordings. A literature study of image processing techniques was conducted, focusing specifically on popular adaptive document binarization methods. During the experimental phase of this study, these methods or parts thereof were applied to the data to determine which techniques would give the most accurate results. Experimentation is the primary research methodology.**

**The iterative experimental phase is discussed in detail as an algorithm is formed to successfully binarize and extract the historic cosmic ray data as well as the temperature of the electrometer while recording. The study concludes with an interpretation of the results obtained in the experimental phase. The success of the algorithm is measured by comparing the resulting data graph to the original.**

**The conclusion of this study is that an adaptive method can be applied to historical recordings of cosmic ray activity to extract numerical data from a wide variety of images without any additional user input.**

# Keywords

- Adaptive Binarization;
- Image Processing;
- Cosmic Rays;
- Digitization;

# Contents

<b>Acknowledgments</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Keywords</b>	<b>v</b>
<b>Table of Contents</b>	<b>vi</b>
<b>List of Figures</b>	<b>xi</b>
<b>1. Introduction</b>	<b>1</b>
1.1. Background . . . . .	1
1.1.1. Cosmic Rays . . . . .	1
1.1.2. GLEs . . . . .	3
1.1.3. Historical Cosmic Ray Data . . . . .	3
1.2. Problem definition . . . . .	5
1.3. Research goals . . . . .	7
1.4. Research methodology . . . . .	8
1.4.1. Description . . . . .	8
1.4.2. Literature study . . . . .	8
1.4.3. Experimentation . . . . .	8
1.5. Scope of study . . . . .	9
<b>2. Literature Study</b>	<b>10</b>
2.1. Introduction . . . . .	10
2.2. Image capture . . . . .	12
2.2.1. Preparation of the document . . . . .	13
2.2.2. Balance between quality and resolution . . . . .	13

## Contents

2.3. Image improvement . . . . .	14
2.3.1. Spatial domain . . . . .	15
2.3.1.1. Intensity transformation functions . . . . .	16
2.3.1.2. Spatial Filtering . . . . .	19
2.3.1.2.1. Smoothing spatial filters . . . . .	21
2.3.1.2.2. Order-statistic filters . . . . .	22
2.3.1.2.3. Sharpening filters . . . . .	23
2.3.1.2.4. Unsharp masking and highboost filtering . . . . .	25
2.3.2. Morphological image processing . . . . .	26
2.3.2.1. Set Theory . . . . .	27
2.3.2.2. Erosion . . . . .	29
2.3.2.3. Dilation . . . . .	30
2.3.2.4. Boundary Extraction . . . . .	31
2.3.2.5. Extraction of connected components . . . . .	32
2.3.2.6. Segmentation using morphological watersheds . . . . .	33
2.3.3. Frequency Domain . . . . .	35
2.3.3.1. Lowpass Filters . . . . .	38
2.3.3.2. Highpass filters . . . . .	40
2.3.3.3. Laplacian Filter . . . . .	41
2.3.3.4. Homomorphic filtering . . . . .	42
2.4. Interpretation . . . . .	42
2.4.1. Pattern recognition . . . . .	43
2.4.2. Digitization . . . . .	43
2.5. Adaptive image binarization . . . . .	43
2.5.1. Otsu's method . . . . .	44
2.5.2. Niblack's method . . . . .	44
2.5.3. Sauvola and Pietkäinen's method . . . . .	45
2.5.3.1. Background binarization . . . . .	45
2.5.3.2. Foreground Binarization . . . . .	46
2.5.4. Gatos and colleagues' method . . . . .	46
2.5.4.1. Preprocessing . . . . .	46
2.5.4.2. Rough foreground estimation . . . . .	47
2.5.4.3. Background estimation . . . . .	47
2.5.4.4. Final thresholding . . . . .	48
2.5.4.5. Post-processing . . . . .	49

## Contents

2.5.5. Adaptive water flow binarization . . . . .	49
2.5.5.1. Region of interest extraction . . . . .	51
2.5.5.2. Stroke width measurement . . . . .	51
2.5.5.3. Stopping threshold for rainfall process . . . . .	51
2.5.5.4. Setting the rate of rainfall . . . . .	52
2.5.5.5. Fast algorithm for finding local minima . . . . .	53
2.5.5.6. Blob extraction . . . . .	54
2.5.5.7. Blob classification and final binarization . . . . .	55
2.6. Digitization of cosmic ray recordings in the frequency domain . . . . .	57
2.6.1. Drevin's method . . . . .	57
2.7. Concluding remarks . . . . .	59
<b>3. Experimental process</b>	<b>60</b>
3.1. Introduction . . . . .	60
3.2. Test data . . . . .	60
3.3. Design Process . . . . .	65
3.4. Success . . . . .	68
3.5. Technology/Software . . . . .	69
<b>4. Process description</b>	<b>70</b>
4.1. Introduction . . . . .	70
4.2. Pre-processing . . . . .	72
4.3. Rough data identification . . . . .	74
4.4. Rough data extraction . . . . .	80
4.5. Rough data binarization . . . . .	81
4.6. Accurate data identification . . . . .	86
4.7. Accurate data extraction and binarization . . . . .	87
4.8. Post-processing . . . . .	88
4.9. Process results . . . . .	89
<b>5. Conclusion</b>	<b>93</b>
5.1. Achieved research goals . . . . .	93
5.2. Additional study and possible improvements . . . . .	93
5.2.1. Improved accuracy . . . . .	93
5.2.2. Improved speed . . . . .	97
5.2.3. Adaptive variables . . . . .	97

## Contents

5.2.4. Skew detection . . . . .	97
5.2.5. Complete binarization . . . . .	97
5.3. Concluding remarks . . . . .	98
<b>References</b>	<b>99</b>
<b>A. Method description</b>	<b>103</b>
A.1. Crop . . . . .	103
A.2. Line . . . . .	104
A.3. Fill . . . . .	105
A.4. Blur . . . . .	106
A.5. Scan . . . . .	106
A.6. Remove . . . . .	107
A.7. Mark . . . . .	108
A.8. Erase . . . . .	108
A.9. Clean . . . . .	109
A.10.Target . . . . .	110
A.11.Connect . . . . .	111
A.12.Scrub . . . . .	111
A.13.Identify . . . . .	112
A.13.1.Identify1 . . . . .	113
A.13.2.Identify2 . . . . .	113
A.13.3.Identify3 . . . . .	113
A.14.Bind . . . . .	114
A.15.Extract . . . . .	115
A.16.Purify . . . . .	115
A.17.Designate . . . . .	116
A.18.Define . . . . .	116
A.19.Plot . . . . .	118
A.20.Paste . . . . .	119
A.21.Insert . . . . .	120
<b>B. Visual summary of the binarization process</b>	<b>121</b>
B.1. Pre-processing . . . . .	122
B.2. Rough data identification: Iteration 1 . . . . .	123
B.3. Rough data identification: Iteration 2 . . . . .	124

## Contents

B.4. Rough data identification: Iteration 3 . . . . .	125
B.5. Rough data identification: Iteration 4 . . . . .	126
B.6. Rough data identification: Iteration 5 . . . . .	127
B.7. Rough data identification: Iteration 6 . . . . .	128
B.8. Rough data identification output . . . . .	129
B.9. Rough data extraction . . . . .	130
B.10. Rough data binarization . . . . .	132
B.11. Accurate data identification . . . . .	135
B.12. Accurate data extraction . . . . .	137
B.13. Accurate data binarization . . . . .	138
B.14. Post-processing . . . . .	140
<b>C. Letter from the editor</b>	<b>141</b>

# List of Figures

1.1. NWU Space Research Unit neutron monitor recordings . . . . .	2
1.2. Arthur Holly Compton with a model C cosmic ray ionization chamber . .	4
1.3. Lindemann electrometer . . . . .	5
1.4. Photographic data of cosmic ray activity . . . . .	5
2.1. An image (left) and its negative version (right) . . . . .	16
2.2. Application of the log transform to an image . . . . .	17
2.3. Possible transformation curves of the Gamma transformation . . . . .	17
2.4. Application of Gamma transformations . . . . .	18
2.5. Application of a filter mask to an image . . . . .	20
2.6. An example of a 3 x 3 smoothing filter . . . . .	21
2.7. Application of a 15 x 15 smoothing filter . . . . .	21
2.8. Example of a weighted 3 x 3 smoothing filter . . . . .	22
2.9. Difference between median filtering and mean filtering . . . . .	23
2.10. Sobel filters . . . . .	23
2.11. Prewitt filters . . . . .	24
2.12. Roberts filters . . . . .	24
2.13. Laplacian filter . . . . .	24
2.14. Application of a sharpening mask using Sobel operators . . . . .	25
2.15. Application of a sharpening mask using Laplacian operators . . . . .	25
2.16. Results of unsharp masking . . . . .	26
2.17. Highboost Filter . . . . .	26
2.18. Set theory operators . . . . .	28
2.19. Translation and reflection . . . . .	29
2.20. The process of morphological erosion . . . . .	30
2.21. Dilation of text to improve visibility . . . . .	31
2.22. Boundary extraction . . . . .	32
2.23. Extraction and labeling of connected components . . . . .	33

## List of Figures

2.24. A labeled set of connected components representing cosmic ray data . . . . .	33
2.25. Three-dimensional representation of a cosmic ray data line segment . . . . .	34
2.26. Watershed segmentation process . . . . .	36
2.27. Symmetry of a Fourier spectrum . . . . .	37
2.28. Conversion of a spatial image to a Fourier spectrum . . . . .	38
2.29. Ideal lowpass filter with decreasing values of $D_0$ . . . . .	39
2.30. Ringing in a filtered image . . . . .	39
2.31. Gaussian filter result with zero ringing . . . . .	40
2.32. Ideal highpass filter results . . . . .	41
2.33. Gatos <i>et al.</i> foreground estimation . . . . .	47
2.34. Gatos <i>et al.</i> background estimation . . . . .	48
2.35. Binarization of a typed document image . . . . .	50
2.36. Neighborhood for contrast measurement . . . . .	52
2.37. The process of finding local minima . . . . .	53
2.38. Phases of adaptive water flow binarization . . . . .	56
3.1. The original image A . . . . .	63
3.2. The original image B . . . . .	63
3.3. The original image C . . . . .	63
3.4. The original image D . . . . .	63
3.5. The original image F . . . . .	64
3.6. The original image F . . . . .	64
3.7. Results of applying existing binarization methods to cosmic ray data . . . . .	66
4.1. Difference between scanning a blurred image and an unblurred image . . . . .	74
4.2. Results of pre-processing . . . . .	75
4.3. Results of Scan method at different tolerances . . . . .	77
4.4. The effect of a single iteration of the Scan, Remove, Mark, Erase methods . . . . .	78
4.5. Results of six Scan, Remove, Mark and Erase iterations . . . . .	79
4.6. Final output of the rough data identification phase . . . . .	80
4.7. Rough data extraction . . . . .	82
4.8. Thickening effect of the Purify method . . . . .	83
4.9. Results of 4 Scan, Remove, Purify and Designate iterations . . . . .	84
4.10. Result of applying Scan method to output from Figure 4.9 . . . . .	85
4.11. Effect of the Define method . . . . .	86
4.12. The effect of the Plot method on a set of disconnected data line segments . . . . .	88

## *List of Figures*

4.13. Difference between applying the binarization process once and twice . . .	89
4.14. The data extracted from image A . . . . .	90
4.15. The data extracted from image B . . . . .	90
4.16. The data extracted from image C . . . . .	91
4.17. The data extracted from image D . . . . .	91
4.18. The data extracted from image E . . . . .	92
4.19. The data extracted from image F . . . . .	92
5.1. Results fitted over original image A . . . . .	94
5.2. Results fitted over original image B . . . . .	94
5.3. Results fitted over original image C . . . . .	95
5.4. Results fitted over original image D . . . . .	95
5.5. Results fitted over original image E . . . . .	96
5.6. Results fitted over original image F . . . . .	96