

1. Introduction

1.1. Background

1.1.1. Cosmic Rays

Cosmic rays consist of about 89% hydrogen nuclei (protons), 10% helium nuclei and 1% nuclei of heavier elements (Simpson, 1983). These particles are discharged from sources such as supernovas, pulsars and neutron stars. The activity of these highly energetic particles is primarily recorded by satellites in Earth orbit, because these nuclei are destroyed upon collision with the atmosphere. However, these collisions cause a discharge of secondary particles in the atmosphere. These secondary particles can be detected by ground-based recording devices, as the ratio between the primary particle activity above the atmosphere and the secondary particle activity below the atmosphere always remains constant (Anchordoqui *et al.*, 2003). These recordings are taken by neutron monitors (Simpson, 2000).

Cosmic rays affect the earth in the following ways:

- Alters the states of elements in electronic integrated circuits, causing transient errors or data loss in electronic devices, especially at high altitudes (Ziegler *et al.*, 1996);
- Contributing factor to the origination of lightning (Gurevich & Zybin, 2005);
- May influence climate changes (Realclimate, 2006);
- Contributing to human exposure to radioactivity (ARPANSA, 2012).
- Contributing to approximately a third of the earth's natural radioactivity;

1. Introduction

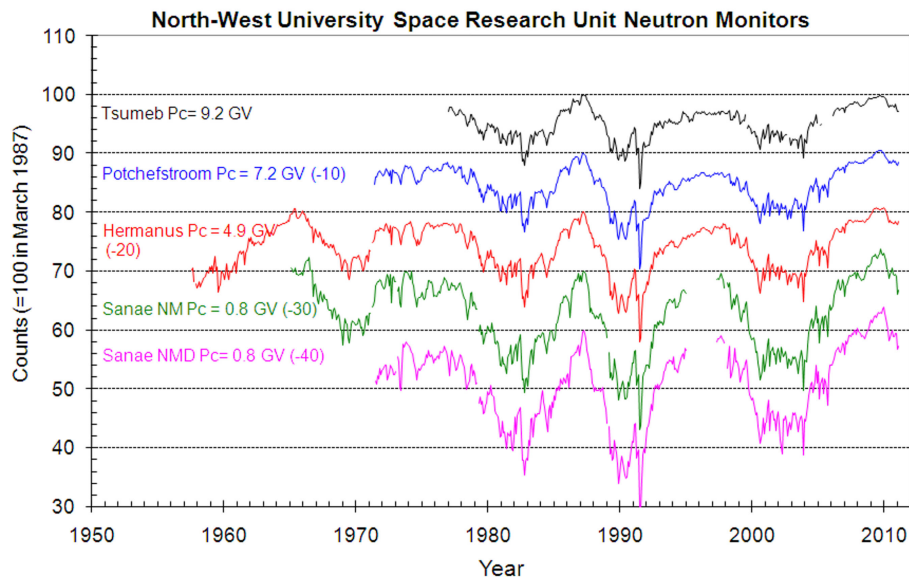


Figure 1.1.: NWU Space Research Unit neutron monitor recordings. (NWU, 2009)

The Space Research Unit of the North-West University has four neutron monitors situated at Sanae, Antarctica; Hermanus, South Africa; Potchefstroom, South Africa and Tsumeb, Namibia (NWU, 2009). Figure 1.1 depicts the recordings of these monitors since these were installed. It is clear that variations in cosmic ray activity appear in all intensities and durations and that cosmic ray activity remains constant across the wide area that these monitors cover.

Maximum intensities of cosmic ray activity occurred in 1965, 1976, 1987 and 2009. These maxima coincide with the solar activity minima. This proves that cosmic rays are not emitted from the sun. However, the sun does affect cosmic ray activity by preventing the rays from penetrating the inner heliosphere when it is more active. This process is known as heliospheric modulation of galactic cosmic rays.

1.1.2. GLEs

The sun, being a typical star, does experience events that may cause highly energetic cosmic rays to be emitted from it. These anomalies usually occur during events such as solar flares. Cosmic rays that are emitted during these events are known as Solar Energetic Particles or SEPs. These SEPs can be so energetic as to be able to penetrate the atmosphere along with the usual cosmic ray particles and these fluctuations in cosmic ray activity can then be recorded at ground level by a neutron monitor, which is why this anomaly is known as a Ground Level Event or GLE. The duration of a GLE is usually only a few hours, making it easily distinguishable from relatively slow changes in normal cosmic ray activity. A GLE is a rare event. Only 70 GLEs have been recorded since 1942 which indicates an average occurrence of once per year.

1.1.3. Historical Cosmic Ray Data

Only during the 1950s did the neutron monitor become the standard instrument for recording cosmic ray activity. The Hermanus neutron monitor is the second longest continuously functioning neutron monitor in the world. Neutron monitors count cosmic ray particles individually and these values are electronically processed and stored, thus these neutron monitors require very little attention and maintenance.

Before the neutron monitors became the standard, cosmic ray activity was mostly recorded using ionization chambers. In the 1930s, the Carnegie Institute in Washington DC gave a number of researchers the task of examining cosmic rays. It quickly became apparent that a worldwide representation of cosmic ray activity would be required to effectively study cosmic rays and their effects on the earth (Lange & Forbush, 1948).

This information led to the placement of seven cosmic ray recording stations which used the Compton-Wollan-Bennett ionization chamber, also known as the model C cosmic ray ionization chamber. Arthur Holly Compton contributed greatly to the design of the device. Figure 1.2 shows him with a model C cosmic ray ionization chamber. The electrometer which was housed inside the chamber is shown in Figure 1.3.

1. Introduction



Figure 1.2.: Arthur Holly Compton with a model C cosmic ray ionization chamber (American Institute of Physics/Science photo library).

The model C cosmic ray ionization chamber used a Lindemann electrometer to measure the ionization of a sphere of purified argon. This measurement was indicated by a needle. The shadow of this needle was projected onto photographic paper, creating graphs of cosmic ray activity. Every hour the chamber was grounded for 3 minutes to reset the electrometer needle to its zero position and also to create a vertical line on the photographic paper to indicate the end of one hour and the beginning of another (Bennett *et al.*, 1934).

Five of these ionization chambers were installed in Maryland, Peru, Mexico, New Zealand and Greenland between 1935 and 1938 while two others were used for specific research goals in Virginia and Colorado in the USA.

In total approximately 114 station years of cosmic ray activity data was recorded and stored on strips of photographic paper. Figure 1.4 is an example of such a photographic data strip. The total length of these strips, if laid end to end, would be longer than 25km.

1. Introduction

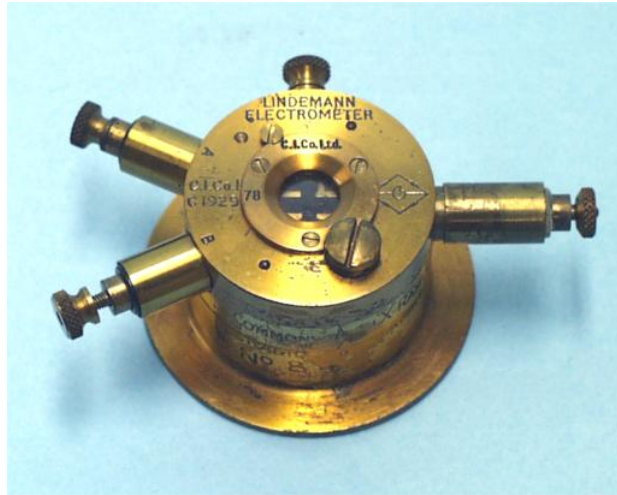


Figure 1.3.: Lindemann electrometer.

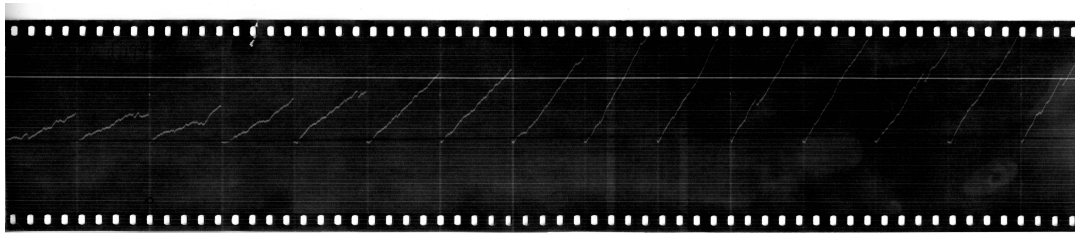


Figure 1.4.: Photographic data of cosmic ray activity recorded by a model C ionization chamber.

1.2. Problem definition

This vast amount of data needs to be accurately digitized and stored in order to be of use for future cosmic ray research. Digitization of the data images requires that the graphs of the electrometer readings be accurately identified and extracted from the background of the image. These extracted graphs then need to be processed into numerical data.

The data was recorded using photographic methods and as a result a variety of photographic and visual defects occur. The presence of these defects results in data images that contain many imperfections or entire images that are difficult to digitize.

1. Introduction

Some of the challenges presented by the data set are:

- The average background values of different images vary greatly;
- The background intensity varies in different parts of the same image;
- The thickness of graph lines differs according to the brightness of the image and the gradient of the graph;
- The intensity values of the graph lines vary greatly between images and in different parts of the same image;
- The graph lines are often intersected by the line indicating the temperature of the device while recording;
- Some images have such high average intensity values that they contain random patches that have the same intensity values as the graph lines to be extracted;
- Some images have such low average intensity values that only segments of the graph lines are barely visible, with only a small difference between the intensity values of these segments and the value of the background in which they are found;
- Some images have such high contrast that the scale lines of the graph have the same intensity values as the graph lines;
- The images are noisy, often containing noise pixels that are of the same intensity as the graph lines;
- Random areas of some images were exposed to light during recording, resulting in unwanted high intensity patches in the image. This usually occurs at the edge of the image. These patches pose a difficulty in very dark images, where the patches have higher intensity values than the graph lines;
- The graph lines are occasionally printed between or over the sprocket holes which were used to move the photographic paper along. In the digital images of the original photographic strips, these sprocket holes appear as rounded rectangles and are the brightest parts of each image. The edges of these sprocket holes may be incorrectly recognized as part of a graph line that was printed between or over them;

The challenge lies in creating a method capable of extracting the data despite the difference in lighting quality and uniformity between different images and parts of the same image. The method should also avoid extracting the variety of unwanted objects with nearly the same intensity value as the graph line. The amount of data also necessitates an adaptive method where the only required user interaction would be the input of the data images.

These requirements can all be met by a collection of image processing techniques.

1.3. Research goals

The primary goal of this research is to develop a fully adaptive image binarization method to binarize and extract data graphs of cosmic ray recordings from digital images of photographic strips that are several decades old.

This method must be able to extract any data found in an image. If the image is of such poor quality that only mere segments of the graph line can be identified, then the method must be able to extract those segments and reconstruct the data graph as much as possible.

The output of this method must fit the graph of the original data accurately, without indications of outliers caused by unwanted objects in the original image.

After adaptability, the second priority of this method will be accuracy, the third being speed.

If this study is successful, then this historical data can be made available to future studies in the field of cosmic ray research. The method could possibly also be applied to other degraded historical data graphs.

1.4. Research methodology

1.4.1. Description

A quantitative approach will be followed. Burns and Grove (1993) define quantitative research as a formal, objective, systematic process to describe and test relationships and examine cause and effect interactions among variables (Burns & Grove, 1993). A literature study will be conducted to gain the required background knowledge to be able to design the methods. The methods will be developed iteratively through experimentation.

1.4.2. Literature study

The literature study will start off with a look into image processing techniques that may be of use in this study. It will also include research on the most popular document binarization methods, to identify key concepts and techniques for the identification and extraction of foreground objects. A brief summary of other studies concerning the extraction of data from cosmic ray recordings is also included.

1.4.3. Experimentation

Techniques and methods identified during the literature study will be applied to the data to see which methods and techniques overcome the different challenges mentioned earlier. The best combination of these methods and techniques will then be derived through experimentation. The optimal values for all parameters will also be determined through experimentation.

This method will be deemed successful when its results can be visually matched to the original image by matching the extracted graph to the thicker graph lines of the original image within an acceptable margin of error, visually compared to the original image. The extracted graph lines should have the width of a single pixel, which should easily fit inside the slightly blurred data lines in the original image. If the extracted lines follow the same pattern and gradient as the original lines by being contained within them, the accuracy of this method will be proved.

This visual measure of success has to be used because there is no data to mathematically or statistically compare the output of this method to. The only possible comparison would be against hand traced lines of the data lines in the original image. Drawing of these lines would rely on the same visual process that is used when comparing results to the original image itself and would not provide a better indication of whether the method is successful or not.

1.5. Scope of study

This study focuses on the binarization and extraction of data from historical and/or degraded graph images. The interpretation of the extracted numerical data falls outside the scope of this study.

Although legacy ionization chamber cosmic ray recordings are used as the data set in this study, the resulting method may be successfully applied to similar degraded graph images.