

Adaptive frequency domain filtering of legacy cosmic ray recordings

GR Drevin

North-West University, Potchefstroom, South Africa

Abstract

Early continuous recordings of cosmic rays, as measured with Carnegie Type C Ionization Chambers, were made on rolls of photographic paper. This paper describes an adaptive frequency domain filtering method to enhance the scale and hour lines as well as the data trace prior to their segmentation and extraction.

Keywords: Filtering, Lorentz curve

1. Introduction

The Carnegie Type C Ionization Chambers [1] were designed and built for the purpose of the continuous recording of cosmic rays. The ionization chamber was a steel sphere containing purified argon at a pressure of 50 atmospheres. The argon was ionized as cosmic rays passed through the chamber. A Lindemann electrometer was used to measure the ionization current. To record the ionization level, the shadow of the electrometer needle was projected onto a moving strip of photographic paper. The barometric pressure and the temperature of the cosmic-ray meter could also be recorded on the same strip of photographic paper. Every hour the ionization chamber was grounded for 3 minutes, zeroing the ionization current and therefore bringing the electrometer needle back to the zero position. At the same time the lamp of the recorder was dimmed resulting in hourly vertical

lines. An example recording is shown in Figure 1.

This data is important in studies of the manner in which cosmic rays respond to changes in the level of solar activity, such as the steady increase that was experienced during the first half of the 20th century.

2. Methodology

The photographic paper recording made at Cheltenham, Maryland, during the ground level event of 19 November 1949 was scanned at a resolution of 600 dpi and 8 bits per pixel using an Epson Perfection 4490 Photo scanner. In a preprocessing step a mask is determined, from the scanned image, to mask out the sprocket holes in the recording as well as the white background at the top and bottom of the image.

An adaptive method for the filtering of the scanned image in the frequency domain is proposed in this paper. Use is made of the Lorentz curve [2] to determine a cut-off frequency for a low-pass filter. The one-dimensional Fourier transforms of the rows and the columns of the image are filtered individually. The result of filtering the rows of the image is used to extract the hour markers in the image while the result of filtering the columns of the image is used to extract the scale lines.

The sub image of each hour is then extracted and processed further. The positions, extents and orientations of the hour

markers as well as the scale lines are obtained using the Hough transform.

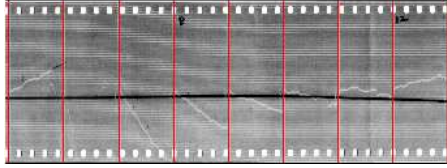


Fig. 1: Part of the recording made on 19 November 1949 at Cheltenham, Maryland. The white irregular line is the ionization current and the thick black line is the barometric pressure. The horizontal white lines are scale lines while the vertical white lines are hour markers.

3. Lorentz curve thresholding

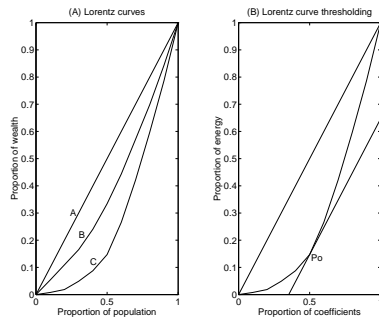


Fig. 2: The Lorentz curve. (A) Example of Lorentz curves. (B) Lorentz curve thresholding.

The Lorentz curve was introduced by Lorentz [2] to measure inequality of income.

3.1. Definition of the Lorentz curve

A population of n individuals with $w(i)$ the wealth of individual $i, i = 1, \dots, n$, sorted

from poorest to richest, is given. Let

$$S(0) = 0$$

$$S(k) = \sum_{i=1}^k w(i), \quad (1)$$

with $S(k)$ the total wealth of the k poorest individuals in the population. Plotting the points

$$(k/n, S(k)/S(n)), k = 0, 1, 2, \dots, n \quad (2)$$

will result in either the straight line

$$y = x, \quad 0 \leq x \leq 1 \quad (3)$$

or a convex curve below the diagonal line defined by Eq. (3) as shown in figure 2(A). The straight line will result if the wealth of the population is distributed equally (line A in figure 2(A)), otherwise a convex curve will result, with the area between the convex curve and the diagonal giving an indication of the inequality of the distribution. The area between the convex curve and diagonal will increase as the income is concentrated in an ever smaller part of the population. In figure 2(A) line C represents a population where the wealth is concentrated in fewer individuals than in the population represented by line B.

3.2. Applying the Lorentz curve

The use of the Lorentz curve as an indicator of the energy distribution and disbalance among the coefficients of the wavelet transform of a data set was proposed by Goel and Vidakovic [3]. They replaced the wealth of a population of individuals with the energy of the transform coefficients. Therefore, given n coefficients, sorted in increasing energy order, with $P(u)$ the energy of coefficient $u = 1, \dots, n$, once again let

$$S(0) = 0$$

$$S(k) = \sum_{u=1}^k P(u). \quad (4)$$

Now $S(k)$ is the total power of the k coefficients with the lowest power.

Goel and Vidakovic [3] and Katul and Vidakovic [4] also proposed a thresholding method based on the Lorentz curve of the energy in the wavelet transform. Their method consists of replacing the $p_0 * 100\%$ coefficients with the smallest energy with 0. The proportion p_0 is defined as:

$$p_0 = \frac{1}{n} \sum_i \mathbf{1}(d(i)^2 \leq \bar{d}^2) \quad (5)$$

where \bar{d}^2 is the mean of $(d(1)^2, d(2)^2, \dots, d(n)^2)$. $\mathbf{1}(A)$ is the indicator function of a condition A , its value is unity if the condition is satisfied and zero otherwise. It is shown by Goel and Vidakovic [3] that p_0 is the argument where the distance between the Lorentz curve and the diagonal is a maximum, given the condition that the coefficients are strictly monotonely increasing. Furthermore, p_0 is the point at which the decrease in coefficients is smaller than the loss in energy. Both the decrease in coefficients as well as the loss in energy are measured on a scale of 0 – 1 and are equally weighted (Figure 2(B)).

In this paper it is proposed that the Lorentz curve be used to determine a cut-off frequency for a filter that is to be used to filter the Fourier transform of images of cosmic ray recordings. To do this $S(k)$ is calculated with the coefficients in decreasing frequency order. The coefficients are kept in frequency order as a filter is not value dependent but position dependent. Now $S(k)$ is the total power of the k highest frequencies and the plot of

$$(k/n, S_k/S_n), k = 0, 1, 2, \dots, n \quad (6)$$

is the Lorentz curve of the energy distribution of the transform. If the energy is uniformly distributed over all the coefficients the Lorentz curve will be the straight line of Eq. 3. As the energy is concentrated in

fewer and fewer coefficients the area between the Lorentz curve and the diagonal will increase.

Normally the energy (power) of a Fourier transform is concentrated in the lower frequencies, with a small fraction of the power being distributed over the high frequency coefficients. The Lorentz curve of the power spectrum of the Fourier transform is calculated and the point where the distance between the Lorentz curve and the diagonal is a maximum is used as the cut-off frequency. It should be noted that this value cannot be assumed to be equal to p_0 as the energy values of the coefficients do not necessarily increase monotonely as their frequencies decrease.

Only the AC components of the Fourier transform are used to calculate the Lorentz curve.

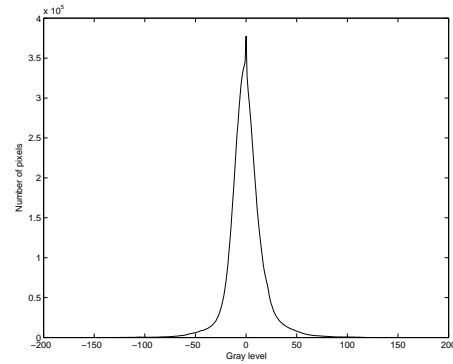


Fig. 3: Distribution of differences between original image and thresholded images.

4. Filtering and thresholding

To filter the image, cut-off frequencies are determined for the one-dimensional Fourier transforms of the rows as well as for the columns of the image.

The Fourier transform of the median of the rows of the image $f(x, y)$ is calculated. Using the Lorentz curve method described

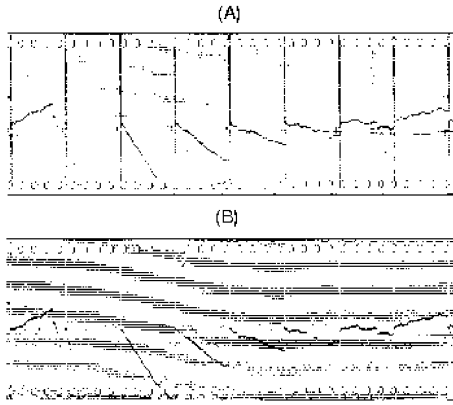


Fig. 4: The filtered and thresholded images. (A) Filtering of Fourier transforms of lines of image and (B) of columns of image.

above, the cut-off frequency, D_r , is obtained from the power spectrum of the row median's Fourier transform. The Gaussian low-pass filter transfer function

$$H_r(u) = e^{-D^2(u)/2D_r^2} \quad (7)$$

with $D(u)$ the distance from the origin of the Fourier transform is then used to filter the Fourier transforms, $F_r(u)$, of all the rows of the image $f(x, y)$

$$G_r(u) = F_r(u)H_r(u). \quad (8)$$

The inverse Fourier transforms of the $G_r(u)$ result in the image $g_r(x, y)$ in which the vertical lines (hour lines) have been smoothed. This image is then subtracted from the original image

$$g'_r(x, y) = f(x, y) - g_r(x, y). \quad (9)$$

The distribution of the values of $g'_r(x, y)$ is shown in figure 3. The pixels of $g'_r(x, y)$ that have positive values are the pixels that have higher values in the original image than in the smoothed image. These pixels, therefore, occur mainly in the vertical lines that have been smoothed.

The image $g'_r(x, y)$ is then thresholded using as threshold the first gray level i where the slope $s(i)$,

$$s(i) = N(i+1) - N(i) \quad (10)$$

is equal to -1. $N(i)$, $i = 0, 1, 2, \dots, P$, in Eq. (10), is the number of pixels, normalized to the interval $[0, P]$ and P is the maximum positive gray level in $g'_r(x, y)$. $N(i)$ has also been smoothed using a moving median of width 7. The result of this thresholding is the image $k_r(x, y)$ which is shown in figure 4(A).

The procedure is repeated for the columns of the image resulting in the thresholded image $k_c(x, y)$ shown in figure 4(B). The distribution of the values of $g'_c(x, y)$, the difference between the original image and the image in which the columns have been filtered, is similar to figure 3.

Figure 7(A) is a part, covering slightly more than an hour, of the logical OR of images $k_r(x, y)$ and $k_c(x, y)$. All of the logical OR of images $k_r(x, y)$ and $k_c(x, y)$ is shown in figure 5

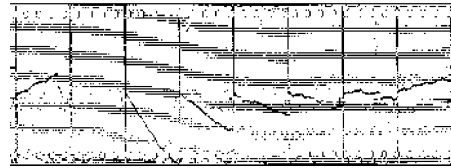


Fig. 5: The filtered and thresholded images.

5. Locating hour markers

Following the thresholding of the image, the hour markers have to be identified. This is done using a simplification of the method proposed by Drevin *et al* [5]. The sum, $h(y)$, of each column of the thresholded image $k_r(x, y)$, that is, the number of pixels in column (y) that are equal to 1 is calculated (figure 6). $h(y)$ is then

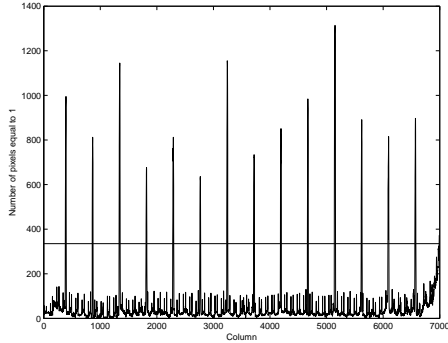


Fig. 6: The sum, $h(y)$, of the columns of the image used to identify the hour markers. The horizontal line is the threshold used in Eq. (11).

thresholded using the threshold proposed by Niblack [6, page 115]:

$$g(y) = \begin{cases} 1 & \text{if } h(y) > \mu + k\sigma \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

with μ the mean and σ the standard deviation of the non-zero values of $h(y)$ and $k = 2$.

After thresholding, all small gaps that may exist in $g(y)$ are removed using

$$g(y) = \text{median}(W_9(y))$$

with $W_9(y)$ a sub array of g with a width of 9 and centered at y . Each of the hour markers should ideally now be represented in $g(y)$ by a single continuous string of ones.

Figure 6 shows $h(y)$ as well as the threshold used in Eq. (11).

To determine the positions, extents and orientations of hour markers and scale lines an individual hour is extracted from the thresholded images using the centers of the hour markers as determined in this step. A 10% overlap is added at the beginning and end of the sub image.

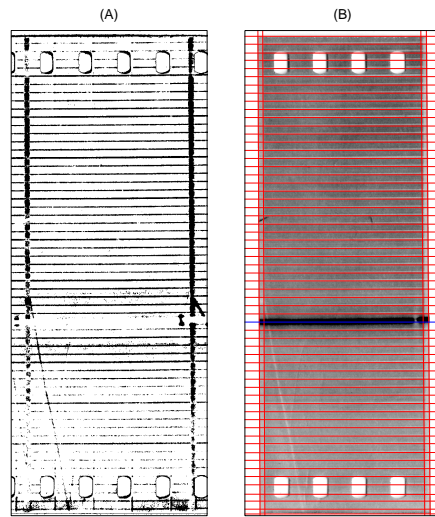


Fig. 7: Image of a single hour. (A) The logical OR of the two thresholded images $k_r(x,y)$ and $k_c(x,y)$. (B) The original image with the edges of the hour markers and the centers of the identified and interpolated scale lines.

5.1. Extracting hour markers and scale lines

The Hough transform ([7]) is used to determine the position, orientation and extent of both the hour markers as well as the scale lines as proposed by Drevin *et al* [5]. In their method the sub image used was extracted from the original image and therefore had to be filtered and thresholded before the Hough transform could be used. They made use of adaptations of the methods proposed by Gathos *et al* [8] [9] and Sauvola and Pietikäinen [10].

Figure 7(B) shows the edges of the hour markers as well as the scale lines for an hour's sub image. A missing scale line that has been identified and interpolated is shown as well.

6. Conclusion and further work

Upon visual inspection of the results it seems that the procedure works well. The procedure however needs to be tested on images of more recordings. Furthermore, synthetic images that are similar to the cosmic ray recordings need to be created to test the procedure.

The procedure now has to be extended to extract the ionization current traces. Once these traces have been extracted they can be converted to numerical values and corrected for geometric distortion using the positions and orientations of the hour markers and scale lines.

References

- [1] A H Compton, E O Wollan, and R D Bennett. A precision recording cosmic-ray meter. *The Review of Scientific Instruments*, 5:415–422, December 1934.
- [2] M O Lorentz. Methods of measuring the concentration of wealth. *Journal of the American Statistical Association*, 9:209–219, June 1905.
- [3] P Goel and B Vidakovic. Wavelet transformations as diversity enhancers. Discussion Paper 95-04, ISDS, Duke University, 1995.
- [4] G Katul and B Vidakovic. The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering. Discussion Paper 95-05, ISDS, Duke University, 1995.
- [5] G R Drevin, H Moraal, and K G McCracken. Determining the skew and scale in images of Compton-Wollan-Bennett ionization chamber recordings. In *Proceedings of the 10th Joint Conference on Information Sciences*, pages 888–894, Salt Lake City, Utah, USA, July 18–24 2007. World Scientific Publishing Co.
- [6] W Niblack. *An Introduction to Digital Image Processing*. Prentice-Hall, 1985.
- [7] P V C Hough. A method and means for recognising complex patterns. U.S. Patent No. 3,069,654, December 18 1962.
- [8] B Gatos, I Praktikakis, and S J Perantonis. An adaptive image binarization technique for low quality historical documents. In S Marinai and A Dengel, editors, *Document Analysis Systems VI*, volume LNCS 3163 of *Lecture Notes in Computer Science*, pages 102–113, Florence, Italy, September 2004. Springer.
- [9] B Gatos, I Praktikakis, and S J Perantonis. Adaptive degraded document image binarization. *Pattern Recognition*, 39:317–327, 2006.
- [10] J Sauvola and Pietikäinen. Adaptive document image binarization. *Pattern Recognition*, 33:225–236, 2000.