5. Conclusion

5.1. Achieved research goals

The primary goal of this research was 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.

The method must be able to accurately extract data lines without being affected by image objects such as sprocket holes, hour markers and scale lines or image characteristics such as varying background intensities or very high/low average intensities.

The proposed method meets all of these requirements, as can be seen when the results are fitted over the original images in Figures 5.1 through 5.6. To improve visibility, only five hours of each resulting image are shown.

This process could also easily be adapted to binarize other degraded/historical graph images.

5.2. Additional study and possible improvements

5.2.1. Improved accuracy

In some images, the extracted data line does not fall exactly within the center of the original data line. The scan method may be improved to avoid this and ensure greater accuracy, without drastically increasing processing time.

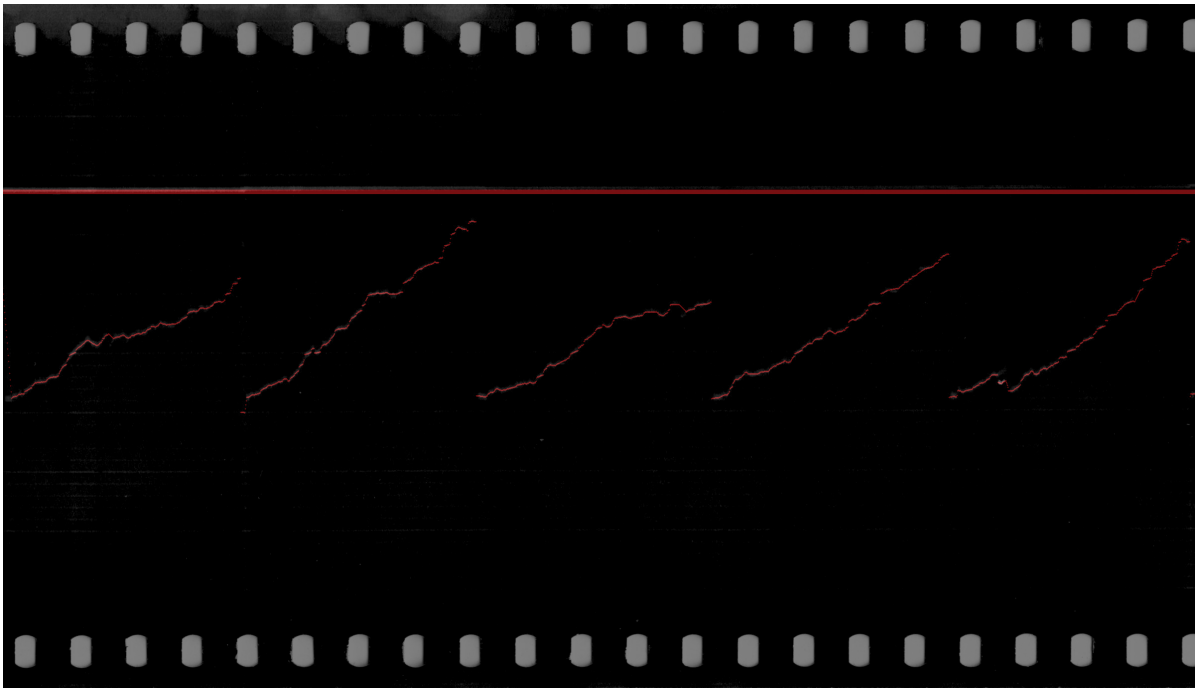

Figure 5.1.: Results fitted over original image A.

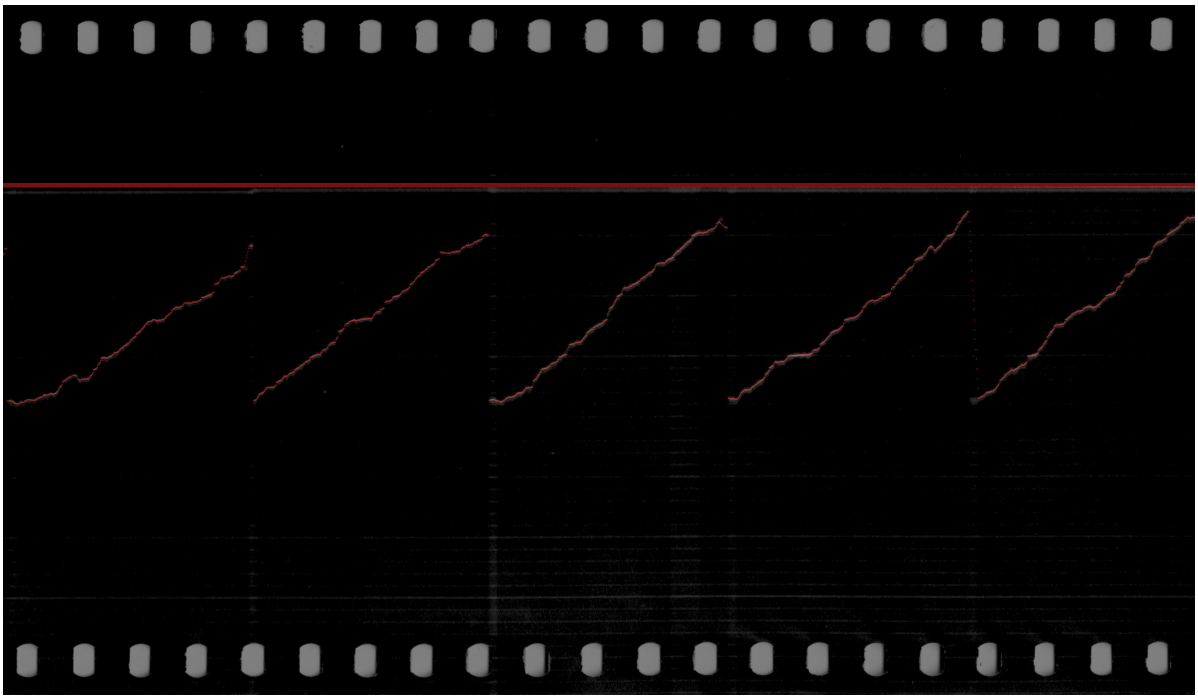

Figure 5.2.: Results fitted over original image B.

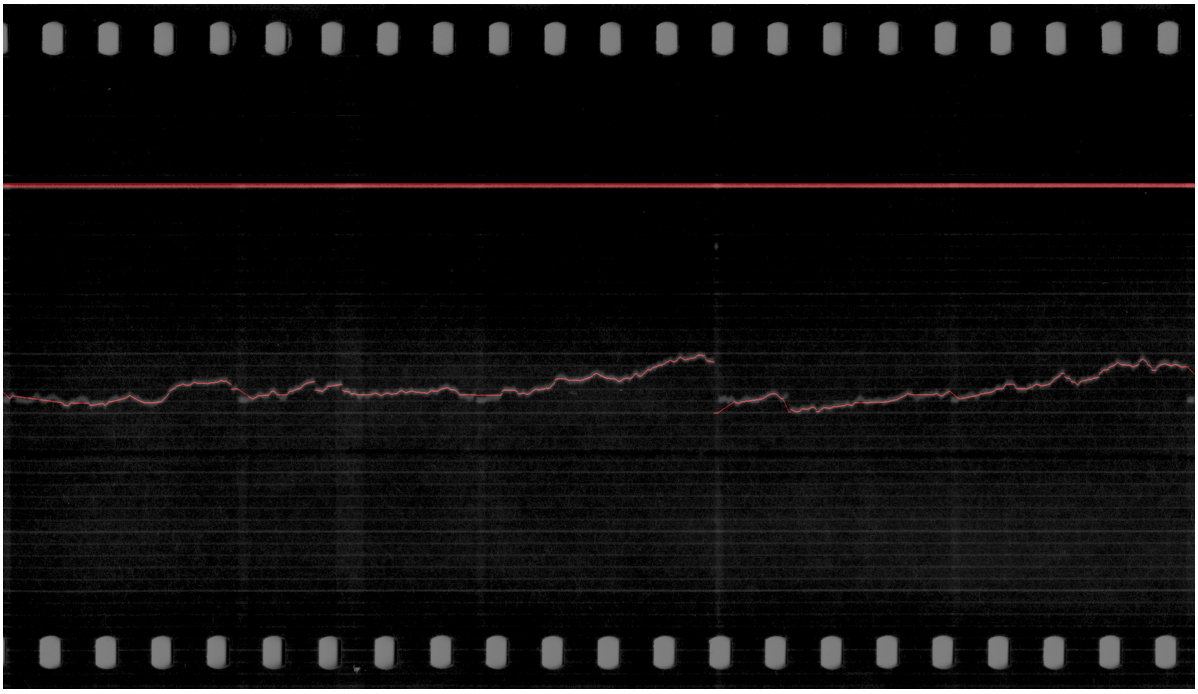

Figure 5.3.: Results fitted over original image C.

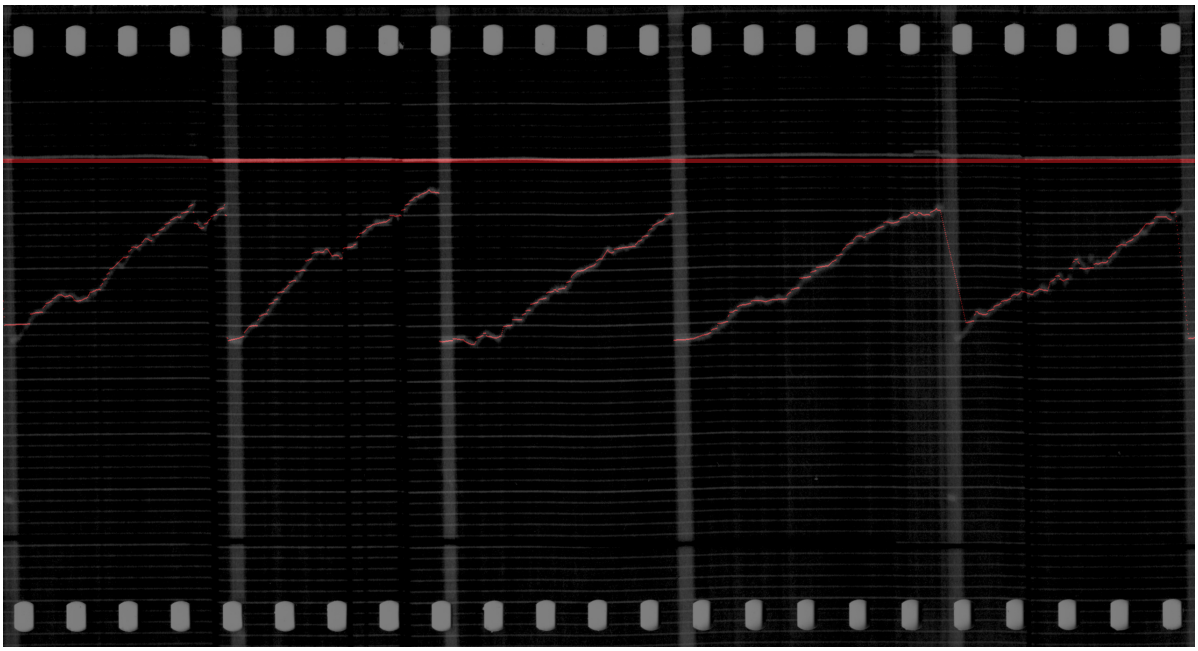

Figure 5.4.: Results fitted over original image D.

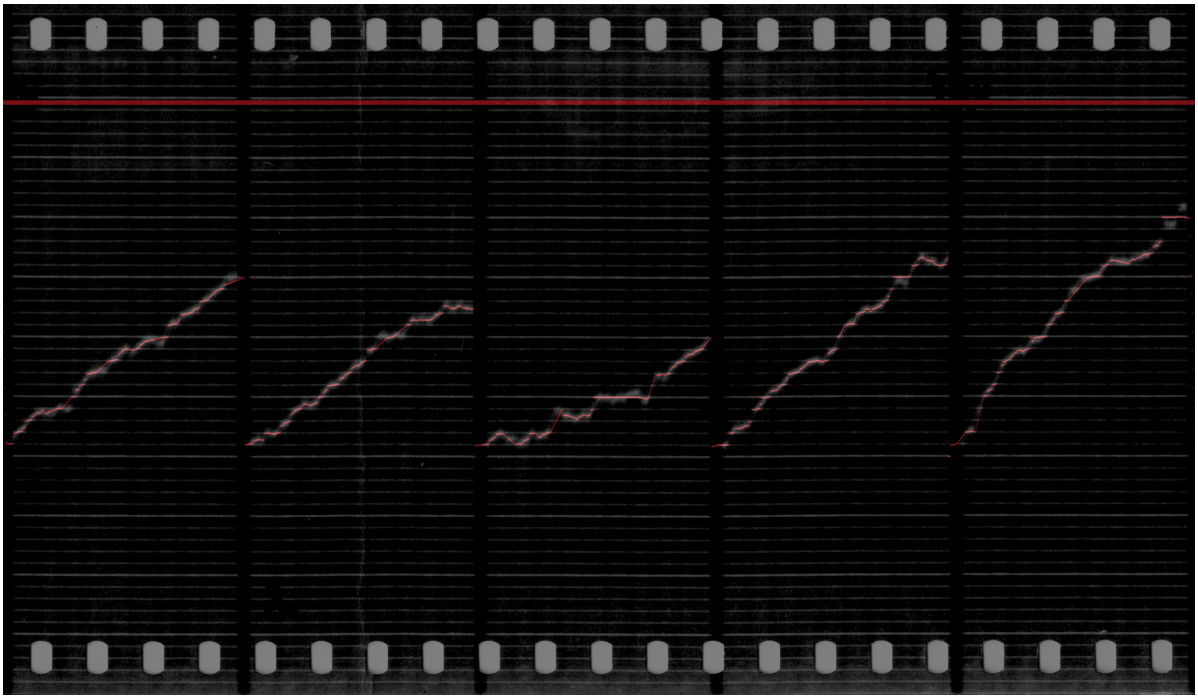

Figure 5.5.: Results fitted over original image E.

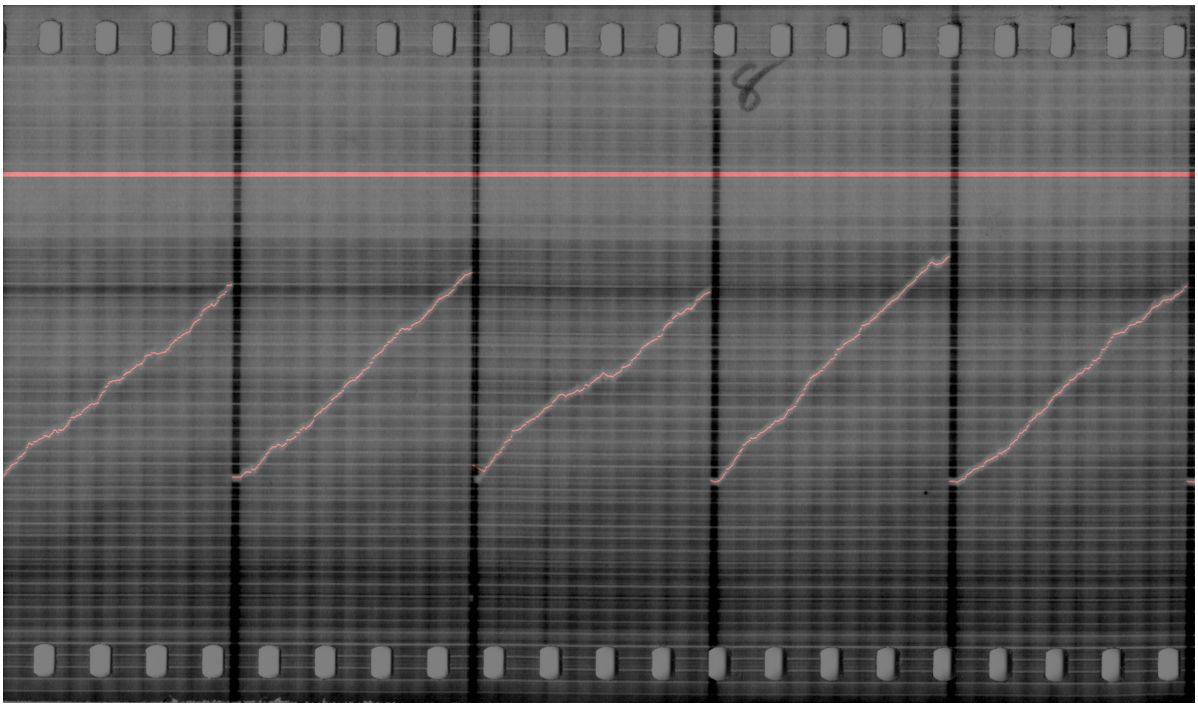

Figure 5.6.: Results fitted over original image F.

5.2.2. Improved speed

The methods used were written and tested in MATLAB 2012. Although MATLAB 2012 is ideal for experimentation and testing, it is much slower than other coding languages. Rewriting the process in one of the C languages or Java could drastically improve performance. This will allow the method to be compiled once and applied to images instead of having the code interpreted as when using MATLAB 2012.

5.2.3. Adaptive variables

The variables such as Scan tolerance, Mark sensitivity and all buffer values used in the Identify methods were determined through experimentation. These values were chosen to provide the best results in their methods while not negatively affecting the results of other methods. One such occurrence is when the sensitivity of the Mark method is decreased to such a degree that all horizontal lines are easily marked as well as several rows containing a large number of pixels belonging to a relatively straight data line.

Such occurrences may be avoided and the accuracy of the method improved if a method was designed to adaptively alter these variables according to image characteristics such as average intensity.

5.2.4. Skew detection

If an accurate skew detection could be applied to the image and all skew removed, then horizontal lines could be detected and removed more accurately. This could reduce the number of iterations of certain methods.

5.2.5. Complete binarization

If this process, which is focused in the spatial domain, is combined with methods applied in the frequency domain, such as those mentioned by Du Plessis (2010), then a process could be created to accurately binarize all foreground objects within a cosmic ray image. The methods used in this study could be applied to extract the highly variable cosmic ray and temperature data, while

the frequency domain methods could be used to extract more constant image objects such as hour markers and scale lines.

5.3. Concluding remarks

This study can be concluded by referring to Figures 5.1 through 5.6 once again. The process can be applied to successfully extract cosmic ray data from a set of images that vary in many ways. The greatest challenge of this study was to balance out the methods and their parameters in such a way that the solution to one problem does not cause many more problems in subsequent phases of the process.