DEVELOPMENT OF SOFTWARE TO DIGITIZE HISTORIC HARDCOPY SEISMOGRAMS FROM NUCLEAR EXPLOSIONS

Steven R. Taylor[1] and Xiaoning (David) Yang[2]

Rocky Mountain Geophysics, LLC[1] and Los Alamos National Laboratory[2]

ABSTRACT

One main issue associated with nuclear explosion monitoring research is the paucity of seismic recordings to be used for calibration purposes. The problem is particularly acute for nuclear explosions that occurred in the pre-digital seismic era. Many of these historic explosions have been detonated under a wide variety of geologic and emplacement conditions and it is important to revisit them in the context of broad area monitoring. Our Phase I work has demonstrated the feasibility of developing software for the digitization of historic hardcopy seismograms from nuclear explosions. We have already begun development of a prototype MATLAB® based Graphical User Interface (GUI) computer code called *DigiSeis* that is specifically designed to meet this goal. Importantly, our project is conducted by seismologists who have a long history of nuclear explosion monitoring with intimate understanding of the nuances of the signal aspects that are important to this field.

The key element of our approach is to use mathematical statistics to automatically determine specified tuning parameters in order to develop spatially adaptive thresholding parameters, accurate timing and calibration information, and signal reconstruction schemes that maintain waveform fidelity. We have implemented some of the statistics algorithms in the current version of *DigiSeis* and have successfully applied *DigiSeis* to digitize some WWSSN LP recordings of NTS explosions. Human interaction with the program is facilitated via a graphical user interface. Scanned photographic images can be read in, cropped and rotated. Image processing techniques are used for adaptive contrast enhancement and horizontal noise trace removal prior to digitization. The core methodology used in *DigiSeis* is global and local histogram enhancement techniques combined with median-value processing. Specialized peak tracking techniques have been developed to isolate high-frequency features such as *P*-wave arrivals where pixel resolution becomes a problem. A Wiggins Interpolator is used to convert pixels to uniform sampling and to fill in signal segments that are not tracked due to various reasons. Digitized seismograms are written to Seismic Analysis Code (SAC) files with timing information estimated through pixel analysis of minute marks.

OBJECTIVES

The technical objective of this study is the development of software for the digitization of historic hardcopy seismograms from nuclear explosions.

RESEARCH ACCOMPLISHED

Our Phase I work has demonstrated the feasibility of developing software for the digitization of historic hardcopy seismograms from nuclear explosions. We have begun development of a prototype MATLAB® based GUI computer code called *DigiSeis* that is specifically designed to meet this goal. Those of us who are old enough to have had the tedious misfortune of having to hand-digitize seismograms, are well aware of the difficulties encountered in dealing with hard-copy seismograms. Importantly, our project is conducted by seismologists who have a long history of nuclear explosion monitoring with intimate understanding of the nuances of the signal aspects that are important to this field. In order to increase the data availability of digital seismic data for nuclear explosion monitoring research it is important to consider systematically scanning and digitizing hardcopy seismograms (both paper and photographic) from historic explosions. Portions (e.g., either *P* waves, *S* waves or surface waves) of particular nuclear explosions have been hand-digitized for specific studies (e.g., Helmberger and Hadley, 1981 as one of many examples). Most available historic seismic data from nuclear explosions comes from the World Wide Standardized Seismographic Network (WWSSN) that began operation in 1961 (e.g., Lee et al., 2005). The SeismoArchives project initiated by W.H.K. Lee is under the auspices of the International Committee for Preservation of WWSSN and Historical Seismograms (ICPWHS) in collaboration with the Incorporated Research Institutions for Seismology (IRIS) Data Management Center (DMC). The original records were photographed and stored on 70 x 120 mm film chips and are available from Lamont-Doherty Earth Observatory and the USGS Albuquerque Seismic Laboratory (ASL). Current scanning programs are underway at the Incorporated Research Institutions for Seismology (IRIS) Data Management Center (DMC; Lee et al., 2005). As discussed by Lee et al., (2005), some early historic nuclear explosion film chip scanning was performed at the ASL for 78 earthquakes and 156 nuclear explosions (about 30,000 film chips were scanned). However, most recent scanning efforts have been focused on significant historic earthquakes.

There is already some existing software for digitizing seismogram images, e.g., SeisDig (Bromirski and Chuang, 2003) and Teseo[2] (Michelini and the Sismos Team, 2005). Nuclear explosions typically have a distinctive character (which is why they can be discriminated from earthquakes) and our goal is to develop a software package that can be tuned to particular signals (including, but not limited to, seismic signals from nuclear explosions). The key element of our approach is to use mathematical statistics to automatically determine specified tuning parameters in order to develop spatially adaptive thresholding parameters, accurate timing and calibration information, and signal reconstruction schemes that maintain waveform fidelity. We have implemented some of the statistics algorithms in the current version of *DigiSeis* and have successfully applied *DigiSeis* to digitize some WWSSN LP recordings of NTS explosions. Human interaction with the program is facilitated via a GUI. Scanned photographic images can be read in, cropped and rotated. Image processing techniques are used for adaptive contrast enhancement and horizontal noise trace removal prior to digitization. The core methodology used in *DigiSeis* is global and local histogram enhancement techniques combined with median-value processing. Specialized peak tracking techniques have been developed to isolate high-frequency features such as *P*-wave arrivals where pixel resolution becomes a problem. A Wiggins Interpolator (Wiggins, 1976) is used to convert pixels to uniform sampling and to fill in signal segments that are not tracked due to various reasons. Digitized seismograms are written to SAC, (http://www.iris.edu/software/sac/) files with timing information estimated through pixel analysis of minute marks.

Because of the large quantity of seismograms that have already been scanned, and the many more that potentially could be scanned, their digitization must be as automated as possible. This is quite challenging because of the varying quality of the photographic records. Figure 1 shows one of the earliest seismograms from a nuclear test illustrating many of the difficulties associated with scanned paper records. For example, it appears that the left portion of the image may have been exposed to sunlight and the background contrasts with the right portion. As will be discussed below, this complicates the preparation of the image for subsequent digitization because background threshold values are spatially dependent. Another phenomenon is that the high-frequency signals have low contrast with the background color, which makes them difficult to track.

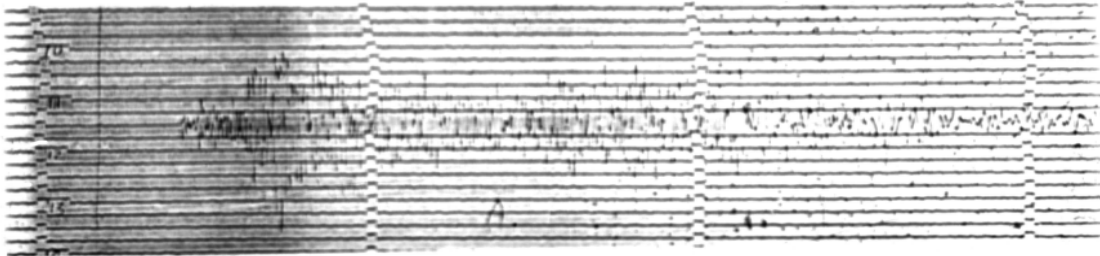## TRINITY
July 16, 1945   recorded at TUC

Figure 1. Low-resolution scan of paper record seismogram recorded at TUC (Tucson, AZ) from the atmospheric TRINITY nuclear test at the Alamogordo Test Range, on the Jornada del Muerto (Journey of Death) desert, NM, on July 16, 1945 (courtesy of Terry Wallace).
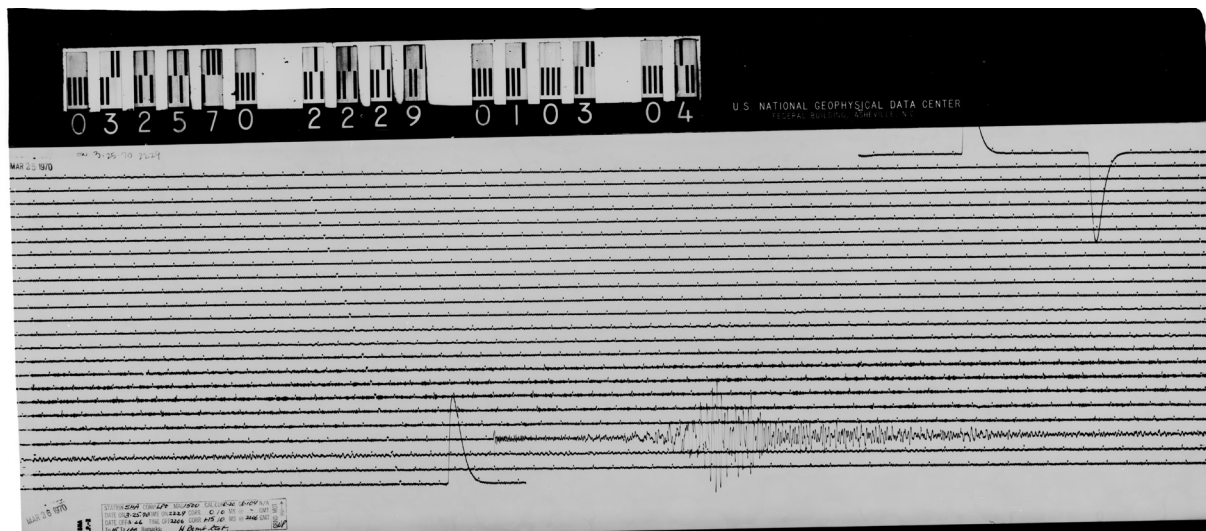

Figure 2. Scanned image of the m_b = 6.5 Handley nuclear explosion at NTS of March 26, 1970 at the long-period WWSSN station SHA (Spring Hill, Alabama; $\Delta \sim 24°$).

As an illustration of our Phase I effort to develop the basic digitization methodology and the prototype software package *DigiSeis*, we use *DigiSeis* to digitize the long-period (LP) WWSSN seismogram of the Handley nuclear explosion and describe the procedure and the results below. Figure 2 shows a scanned seismogram obtained from the IRIS SeismoArchives from the m_b = 6.5 Handley NTS nuclear explosion of March 26, 1970, on the LP vertical component at the WWSSN station SHA (Spring Hill, Alabama; $\Delta \sim 24°$). The seismograms were recorded on a rotating circular drum and consist of a single day's worth of record. The stamp in the lower left-hand corner gives information such as the station name, start and end times, natural period of the seismometer and galvanometer, and magnification (gain). Early in the signal are calibration pulses that can be used to check the amplitude and phase calibration (e.g., Mitchell and Landisman, 1969). Timing marks (ticks) from crystal oscillator clocks mark every minute and wider ticks for each hour. Every sixth hour has no hour mark. The WWSSN 70 x 120 mm film chips were scanned with a resolution of 3200 dpi (Lee et al., 2005).

Figure 3 is a screen shot of the *DigiSeis* GUI with the Handley seismogram displayed. The GUI consists of four figure panels and three groups of function buttons. The *Image* panel displays the original loaded image or the processed image. The *Digitize* panel displays the digitized trace along with the image for comparison. Since no digitization has been done in Figure 3, this panel, as well as the *TraceSelection* and *Threshold* panels, is blank. We will explain in detail the functions of different buttons and other GUI elements below. Once a seismogram image is read into *DigiSeis*, the user can crop the image to include only the portion of the image that contains the signal of interest. If the image is skewed, as is the Handley seismogram shown in Figure 3, the user can rotate the image by selecting two points to define the axis of rotation. These functions are fulfilled by clicking the *Crop* and the *Rotate* buttons. Figure 4 shows the image after cropping and rotation.

There are numerous ways that the seismogram image shown in Figure 4 can be digitized (e.g., Gonzalez and Woods, 2002). Figure 5 is a screen shot of *DigiSeis* illustrating the digitization process. The basic technique that we use for the digitization is called thresholding, in which image pixels are divided based on their grayscale values. We first construct a histogram of the normalized grayscale values from the image shown in the *Image* panel of Figure 3. The histogram is plotted in the *Threshold* panel in Figure 5 where black pixels have a value of 0 and white pixels have a value of 1 (note the gradation indicated by the horizontal bar below the histogram). The histogram shows a single large peak centered at about 0.82 units, which results from the gray image background. The small number of pixels having values of 0.3 and below represent the darker lines associated with the horizontal signal and noise traces. In applying the thresholding, we set a grayscale-value threshold and remove all pixels that have values larger than the threshold. This threshold is indicated by the vertical red line in the *Threshold* panel. After thresholding, we then use the locations of small-valued pixels that survived the thresholding to calculate the locations (row numbers) of the signal trace using a median-value technique, in which we compute the median of pixel grayscale values for each pixel column, find the row location of the corresponding pixel and use this location as the location of the signal trace that we digitize. Using the median, instead of the mean, reduces the effect of extraneous noise that may appear in outlying rows. The user can adjust the threshold either by moving the slider bar, by entering a value, or by dragging the vertical red line. As the threshold value is adjusted, the digitized trace, shown as the red line in the *Digitize* panel, is recalculated and the plot is updated, thus providing an effective method for selecting a proper global threshold value.

Oftentimes the cropped image contains multiple noise traces in addition to the signal trace to be digitized, as Figure 4 shows. We implemented a technique in *DigiSeis* to mask the noise traces (the *SelectTrace* button). The procedure is often necessary because of the median processing technique that we use to track the signal. In *SelectTrace*, we first calculate the sums of grayscale values along the rows of the image shown in Figure 4 and plot the result in the *TraceSelection* panel in the lower left-hand of the *DigiSeis* GUI (Figure 5). In the plot, each peak corresponds to a narrow, horizontal and dark region of the image, such as a signal trace. We use a cutoff value (the blue line) to identify the peaks (traces) in the plot. The user can then select which trace to unmask using the bottom slider bar. Once *TraceSelection* is completed, white masks will appear on the image in *Image* panel (Figure 5) indicating that these horizontal swaths will not be used in the digitization processing. As will be described below, we have been testing an alternative method for removing horizontal noise traces by accurately tracking and masking noise traces themselves rather than masking with horizontal swaths. Our preliminary results indicate that this method may improve the ability to detect signal peaks that are very close to the noise traces.

After trace selection and grayscale-histogram calculation, we choose a global threshold value of 0.7 based on the histogram, which appears to give a good tradeoff between signal resolution and randomly distributed noise pixels for the Handley seismogram. The *Digitize* panel in Figure 5 shows the digitized trace (the red line) superimposed on the original image following the application of the global thresholding.

In studies of nuclear explosions (e.g., Wallace et al., 1983), it is critically important to accurately digitize the *P* waves. Subtle differences in the P waveform contain important information regarding emplacement depth and nature of the explosion source (e.g., Lay, 1991). Figure 5 shows a good tracking of the overall waveform for Handley however; close examination of the P wave illustrates deficiencies that must be remediated through specialized processing. We implemented several pre-processing techniques in *DigiSeis* to enhance image contrast between signal and background pixels (see additional function buttons in the *Image* button group in Figure 3). These techniques become particularly important in cases where a fast moving galvanometer or pen results in a low contrast between the background and the signal. We plot the cropped *P* wave from the Handley explosion in Figure 6 to illustrate the problem.

Figure 3. **A computer screen shot of the *DigiSeis* GUI with the Handley nuclear-explosion seismogram displayed.**
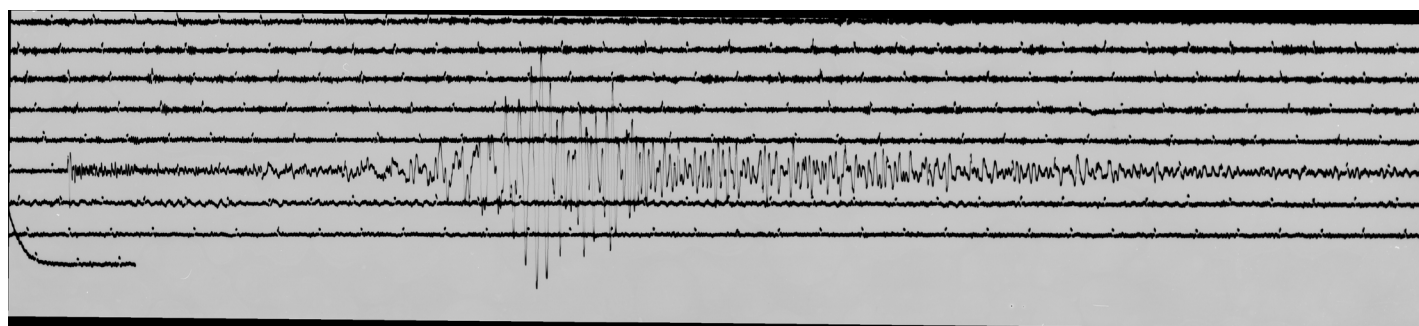

Figure 4. **Cropped and rotated portion of Handley explosion seismogram.**

The left panel of Figure 6 shows an enlargement of the *P* wave from Handley using the global thresholding parameter of 0.7. Although the tracking of the full seismogram in Figure 5 is quite good, it is seen that the global thresholding does not pick up the faint rapid downswing of the *P* wave. Additionally, the tracked first upswing does not track image well by being too gradual. The reason for this is seen by examination of the image histogram in the lower right portion of Figure 5. The broad peak above the assigned threshold of 0.7 represents light gray associated with the image background. Pixel values of the downswing are only slightly less than 0.7 indicating that some contrast enhancement techniques are required. Additionally, the width of the downswing and subsequent upswing is only about 10 pixels.
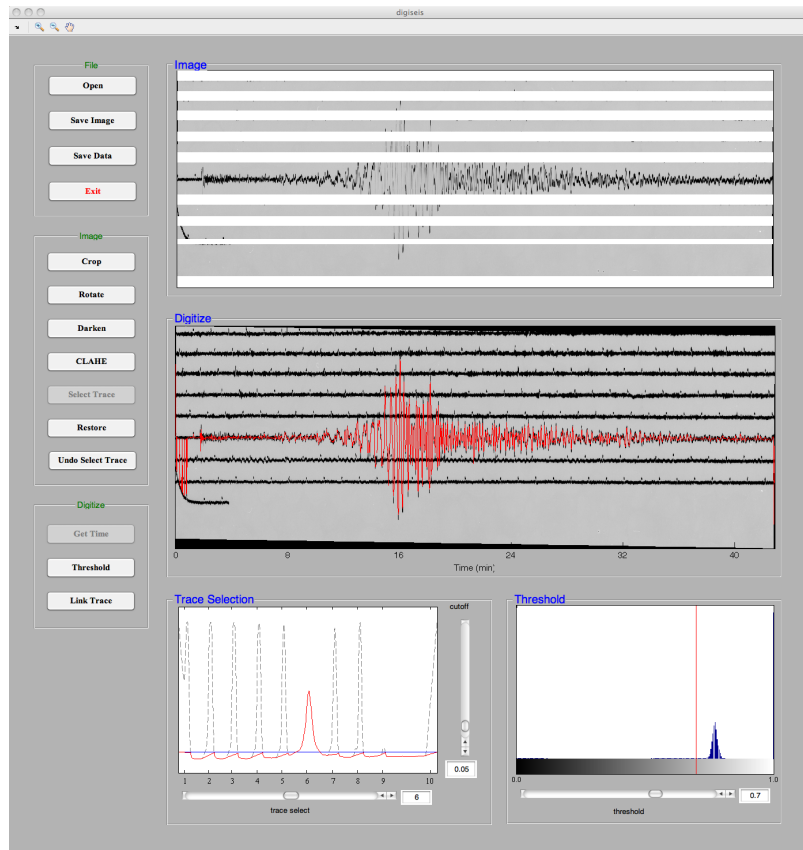
Figure 5. **Screen shot of *DigiSeis* showing various processing steps using the cropped and rotated image shown in Figure 4.**
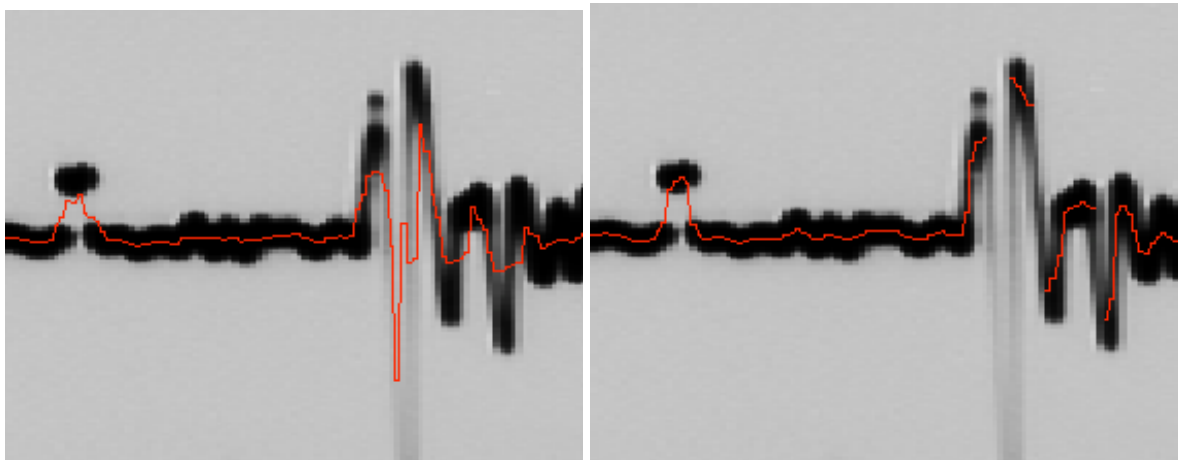

Figure 6. **Close up of Handley *P* wave using global thesholding values of 0.7 (left) and 0.15 (right).**

In the right panel of Figure 6, we have lowered the global threshold parameter down to 0.15. Now the high frequency portions of the waveform are tracked better including the first upswing. However, low contrast features such as the first downswing do not survive the thresholding and are not tracked. This is further illustrated in Figure **7** showing the row and column matrix representation of the image after thresholding. The above example suggests that local processing techniques are required in certain situations to adjust to small nuances associated with specific signal and background characteristics of the image. We have tested a number of methods and find that a *Darken* and *Contrast-Limited Adaptive Histogram Equalization* (CLAHE) appear to give the best results. The *Darken* button can be used to enhance dark pixels associated with signals using a power-law transformation of the form

$$g(x,y) = f(x,y)^{\frac{1}{1+\beta}} \tag{1}$$

where $f(x,y)$ is the input image, $g(x,y)$ is the output image and $-1 < \beta \leq 0$ is a user selectable parameter. Global contrast enhancement uses a whitening transform to make a given image histogram approach a uniform distribution (e.g., Gonzales and Woods, 2002). *CLAHE* is a modification of global contrast enhancement whereby a small region is moved along the image and a local contrast enhancement is performed at each spatial point (e.g., Pizer et al., 1987). Figure 8 shows an example where we have applied *CHAHE* to the Handley P wave. In addition, we have developed a specialized peak finding algorithm (*LinkTrace* button in the *DigiSeis* GUI) that can be used to trace waveforms that are very similar to background. *LinkTrace* first automatically finds the columns of the untracked waveform segments, runs a smoothing operator across each row and at each time (column), and finds the pixel having the minimum value for the corresponding column. Close examination of Figure 8 shows that *LinkTrace* works as desired. We will further show below that if the peaks (troughs) of the signal can be identified, then the Wiggins interpolator algorithm that we use actually can track the rapid swings quite well.
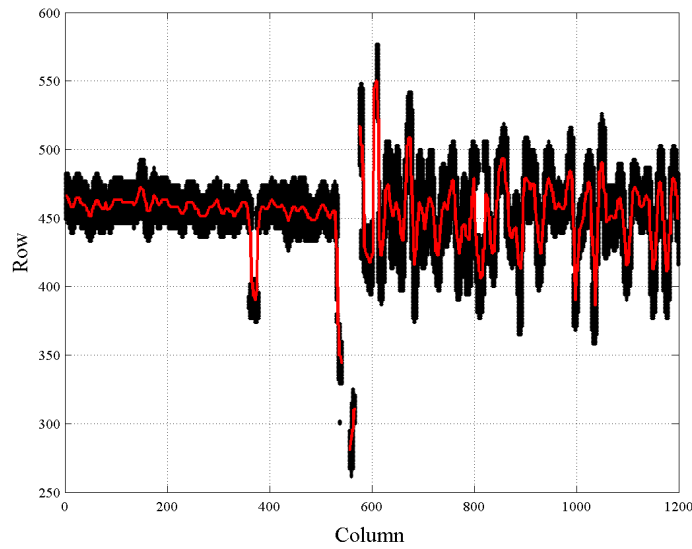

Figure 7. **Row and column matrix representation of the *P* wave in the right panel of Figure 6. Dark corresponds to pixels less than the global threshold parameter of 0.15. Red line corresponds to tracked median row value. Note that because of the matrix storage, the polarity appears reversed.**
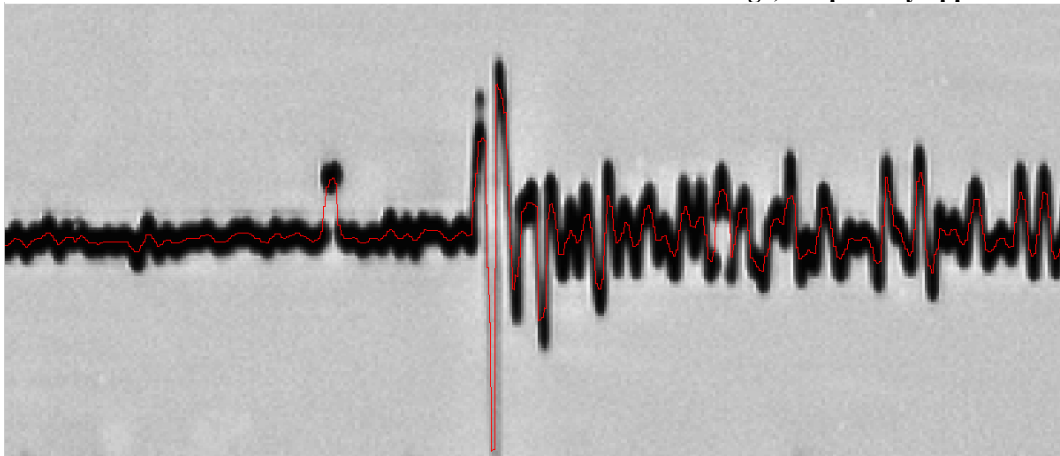

Figure 8. **Handley *P* waveform after application of *CLAHE* and a specialized peak finder (*LinkTrace*) in DigiSeis.**

Another critical aspect of P-waves from nuclear explosions is that m_b measurements need to be accurately estimated both for discrimination (e.g., m_b - M_s) and for yield estimation. Magnitude measurements require a peak-to-peak measure (and sometimes corresponding period) so that the digitized *P* waveform has to be as precise as possible. Figure 9 shows a *DigiSeis* screen shot of the Handley *P* wave and adjacent noise traces. Note that the first downswing barely crosses the underlying noise trace. We have been testing an alternative technique to remove just the adjoining noise traces (see white out lines in the *Image* panel of Figure 9), instead of wide swaths of pixels around the noise traces, in order to retain signal peaks and troughs that are very close to the noise traces. We use the global thresholding to accomplish this. In Figure 9, we have not used *LinkTrace*, and it can be seen that only the later half of the P-wave downswing is tracked. However, we will show below that once a peak or a trough has been identified, our Wiggins interpolation method will capture the other half of the downswing.

Once a signal image has been digitized, it is necessary to output the result into a commonly used seismic data format such as SEED (Standard for the Exchange of Earthquake Data), SAC (Seismic Analysis Code), or CSS (Center for Seismic Studies). Currently, *DigiSeis* outputs SAC files with timing obtained from minute marks (the *SaveData* and *GetTime* buttons). The user uses *SelectTrace* to select a quiet noise trace with clear minute marks to estimate the time scale. The average number of pixels per minute across the trace is computed and subsequently used to convert pixel separation into time. Figure 10 shows the window that pops up when *GetTime* button is clicked. This method of obtaining timing information is still crude and we will develop more accurate techniques for time estimation in the Phase II of the project. The *SaveData* button is used to save the digitized signal. Currently, only SAC file format is supported. Segments of the signal that were not tracked due various reasons are filled with values interpolated from adjacent tracked segments using a Wiggins Interpolator (Wiggins, 1976). The Wiggins interpolator uses a weighted-average-slope method that was specifically designed by seismologists to interpolate hand-digitized seismograms. Figure 11 shows a SAC plot of the digitized Handley seismogram from the image shown in Figure 4. At this point the unit of the signal is in pixels. As part of Phase II project, we will develop methods to derive instrument responses from the calibration pulses and convert the digitized signal to ground displacements or velocity.
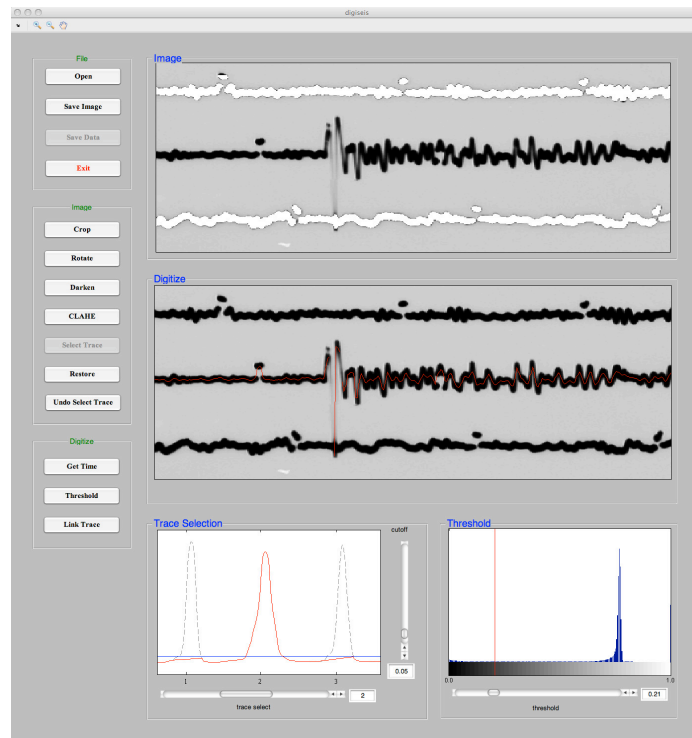

Figure 9. *DigiSeis* **screen shot showing Handley *P* wave and adjacent noise traces. Note that the first downswing barely crosses the underlying noise trace. We have been testing an alternative technique to remove adjoining noise traces (see white out lines in the *Image* panel) to isolate peaks and troughs in the signal, which are very close to the adjacent noise traces. Accurate tracking of *P*-wave amplitudes is important in m_b estimation, which is critical to both discrimination and yield estimation.**
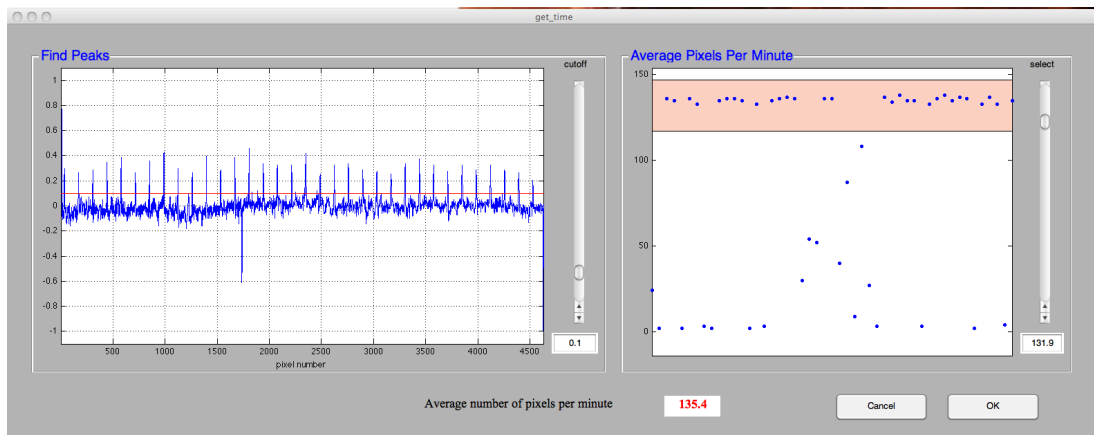
Figure 10. **Screen shot of the window that pops up when *GetTime* button of *DigiSeis* (Figure 3) is clicked. Minute marks manifested as peaks in the left panel are found and average number of pixels between these peaks is calculated (right panel). The average value is then used to convert pixel numbers into time intervals.**
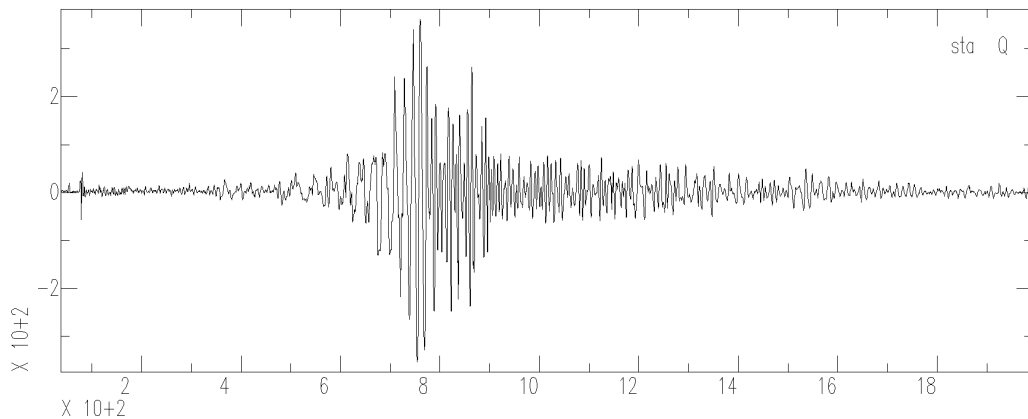

Figure 11. **SAC plot of digitized Handley seismogram (pixels; see Figure 4). The signal is interpolated using the Wiggins interpolator (Wiggins, 1976).**
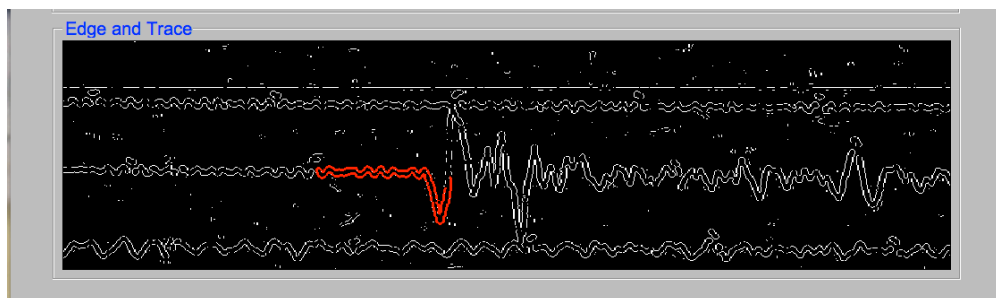

Figure 12. **Boundary of signal edge (red line) superimposed on a *P*-wave image. The white lines correspond to edges detected using the LoG Laplacian gradient operator.**

We have tested an edge detection and tracking algorithm with mixed results. The algorithm takes the Laplacian of a Gaussian (LoG) mapping rapid changes in the image grayscale level and detecting edges. Figure 12 shows a portion of an early version of *DigiSeis* where we had implemented the edge detector. The white lines in the *EdgeAndTrace* panel are the edges defined by the application of the LoG Laplacian gradient edge detector. After the application of LoG, we used an edge tracking and boundary mapping methodology to link the edge boundaries. Figure 12 illustrates that edge detection and boundary tracking is not as robust as the basic global thresholding. For example, the signal on the first main upswing cannot be successfully tracked. Of course the LoG operator can be tuned to better define these rapid swings, but at the expense of increasing noise and false contours.

CONCLUSIONS AND RECOMMENDATIONS

We have developed a prototype MATLAB® based Graphical User Interface (GUI) computer code called *DigiSeis* for the digitization of historic nuclear explosions from hardcopy records.

ACKNOWLEDGEMENTS

We have utilized scanned seismograms from the SeismoArchives scanning project of the Incorporated Research Institutions for Seismology; http://www.iris.washington.edu/data/historical.

REFERENCES

Bromirski P. D. and S. Chuang (2003). SeisDig: Software to digitize scanned analog seismogram images, Users Manual, *Scripps Institution of Oceanography Technical Report*, San Diego, CA.

Gonzalez, R. C. and R. E. Woods (2002). *Digital Image Processing*, Pearson Prentice Hall, Upper Saddle River, NJ.

Helmberger, D. V. and D. M Hadley (1981). Seismic source functions and attenuation from local and teleseismic observations of the NTS events JORUM and HANDLEY, *Bull. Seism. Soc. Am.*, 71: 56–67.

Lay, T. (1991). The teleseismic manifestation of pP: Problems and paradoxes, in *Explosion Source Phenomenology*, *Geophys. Monograph* 65, S. R. Taylor, H. J. Patton and P. G. Richards, eds., 127–140.

Lee, W. H. K., C. R. Hutt, T. K. Ahern, D. Barnes, R.Benson, T. Knight and J. Schweitzer (2005). Constructing SeismoArchives at the IRIS DMC, Appendix to ICSU report.

Michelini, A. and the SISMOS Team (2005). Collection, digitization and distribution of historical seismological data at INGV, *Trans. Am. Geophys. Un*., 86.

Mitchell, B. J. and M. Landisman (1969). Electromagnetic seismograph constants by least-squares inversion, *Bull. Seism. Soc. Am.*, 59: 1335–1348.

Pizer, S. M., E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. H. Romeny, J. B. Zimmerman, and K. Zuiderveld (1987). Adaptive histogram equalization and its variations*, Comp. Vision, Graphics, and Image Proc.*, 39: 355–368.

Wallace, T.C., D.V. Helmberger, G.R. Engen (1983). Evidence of tectonic release from underground nuclear explosions in long-period *P* waves, *Bull. Seism. Soc. Am.* 73: 593–613.

Wiggins, R.A. (1976). Interpolation of digitized curves, *Bull. Seism. Soc. Am.* 66: 2077–2081.