Computers & Geosciences 93 (2016) 70–76

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Research paper

A new program on digitizing analog seismograms

Maofa Wang [a,b,*], Qigang Jiang [a], Qingjie Liu [b], Meng Huang [b]

[a] College of Geoexploration Science and Technology, Jilin University, Changchun, Jilin 130026, China
[b] Department of Disaster Information Engineering, Institute of Disaster Prevention, Sanhe, Hebei 065201, China

ARTICLE INFO

Article history:
Received 26 October 2015
Received in revised form
11 May 2016
Accepted 12 May 2016
Available online 14 May 2016

Keywords:
Analog seismograms
Digitization
Color scene filed Method (CSFM)
Traces mosaic

ABSTRACT

Historical seismograms contain a great variety of useful information which can be used in the study of earthquakes. It is necessary for researchers to digitize analog records and extract the information just as modern computing analysis requires. Firstly, an algorithm based on color scene filed method is presented in order to digitize analog seismograms. Secondly, an interactive software program using C# has been developed to digitize seismogram traces from raster files quickly and accurately. The program can deal with gray-scale images stored in a suitable file format and it offers two different methods: manual digitization and automatic digitization. The test result of the program shows that the methods presented in this paper can lead to good performance.

1. Introduction

Digital seismic stations have been installed in many countries at present, which can provide digital seismograms. However, the analog seismograms have been recorded prior to the digital seismic instrumentation for about 70 years and this large amount of data is still of great value. These original analog seismograms are closely associated with the earthquake-generating process and can be applied in the analysis and study of earthquakes. If these analog seismograms can be digitized, geophysicists will benefit a lot from the results when doing researches on earthquakes (Bungum et al., 2003; Xu et al., 2008).

A great number of analog seismograms have been stored in seismic stations located in countries throughout the world. Sadly, these paper seismograms are piled up in warehouses, buried in dusts, mildewed by moist air, eaten by rats, and suffering from erosions of various natural phenomena (Bungum et al., 2003).

With the development of computer technology, a large number of digital or analog seismograms spanning a longer period are needed in seismic research. Unfortunately, analog seismograms still cannot be processed quickly and accurately by computers in a big-data era even if they remain an important part of seismic records in human history. Without digitization, the precious historical value of these analog seismograms cannot be fully realized (Bogert, 1961).

A project named HRV Seismogram Archival was conducted in Harvard University in mid 2012. The research team had invested a lot of money and effort to clean and rasterize old seismograms in the HRV station by scanning paper analog seismograms (Harvard station, 2015). After being processed, these analog seismograms can be stored and released in the form of raster images, which is more convenient to store and read. However, without further digitization, the raster images still cannot be directly processed by computers. The usual methods applied to rasterizing analog seismograms are photographing or scanning, which prolongs their storage period, but their importance is undervalued. In other words, they only exist but are seldom used.

Today, digitization of analog seismograms is usually based on raster images scanned from paper records (Trifunac et al., 1999). CAD core was used to vectorize analog seismograms (Teves-Costa et al., 1999). A vectoriser named Teseo was developed to digitize historical seismograms, which offers three digital methods: manual, automatic by color selection, and automatic by neural networks (Pintore et al., 2005). Xu and Xu (2014) developed an interactive program on digitizing historical seismograms using matlab based on point tracing algorithm.

Although great improvements have been achieved, results from the automatic digitizing programs still have defects in some particular situations (Trifunac et al., 1999), which need to be corrected subsequently. In this paper, we present an algorithm for automatic digitization of the seismograms based on color scene filed method (CSF), then an interactive program is developed using C# to digitize seismogram traces from raster files quickly and accurately. It can handle gray-scale images stored in a suitable file format and it offers two different methods: manual digitization and automatic digitization.

* Corresponding author at: College of Geoexploration Science and Technology, Jilin University, Changchun, Jilin 130026, China.
E-mail address: wangmaofa2008@126.com (M. Wang).

http://dx.doi.org/10.1016/j.cageo.2016.05.004
0098-3004/

2. Digitization

2.1. Principle

An overall digitization scheme presented in Fig. 1 shows that there are altogether six steps: Seismograms Acquisition, Information Association, and Seismograms to Raster, Seismograms Tracing, Mosaic and Inversion. In the following, a detailed discussion and description of these key steps will be given.

2.2. Seismograms acquisition

In our study, we focused on about 500 seismograms acquired in 1991 stored in the seismic station of Chengde, a small city located in North China. In 1991, several earthquakes of magnitude over 6.5 hit cities in the northern part of China. These paper seismograms were scanned into raster images by professional scanners.

2.3. Information association

In this step, we need to associate the detailed information about the location of seismic stations, instrument parameters, and earthquake bulletins, etc. All these information will be indispensable in the steps of time mosaic and seismogram analysis after seismograms tracing.

2.4. Seismograms to raster

Historical seismograms are often recorded on smoked paper or photographic record, both of which are given in grayscale colors. Professional scanners are usually used to scan analog seismograms. The output of scanners is raster image, which are made up of two-dimensional arrays of pixels. Each pixel is represented by the gray value of the point. Raster images in different types can be converted into gray-scale images in order to unify the color of the images. The resolution of the scanned images must be high enough to retain sufficient information for the later processing of digitization. In this paper, different formats of raster images are firstly converted into PNG format.

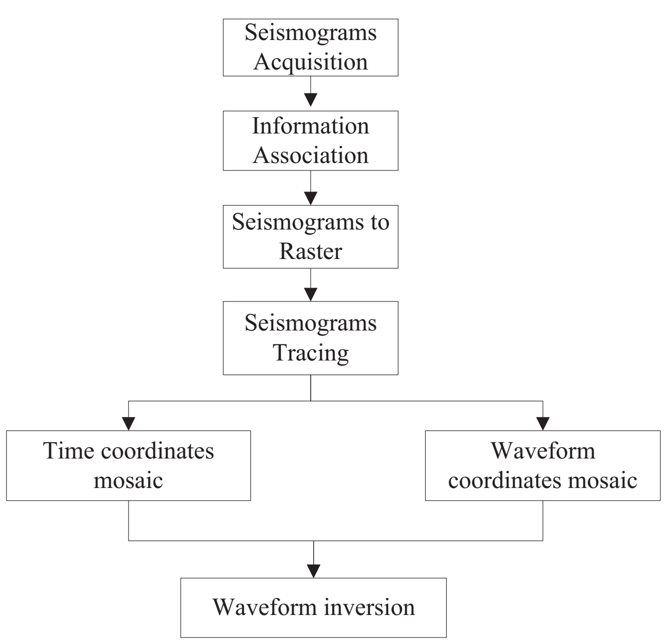

Fig. 1. An overall scheme to digitize analog seismograms.

2.5. Seismograms tracing

2.5.1. Preprocessing

A grayscale raster image is merely a two-dimensional matrix, and each element in the matrix has a value ranging from 0 (pure black) to 255 (pure white) (Xu and Xu, 2014). To reduce the noise and complexity of the following tracing, image binaryzation is required. Binaryzation is a technique of digital image processing, by which one image can be divided into two parts: the object part and the background part. Otsu algorithm is used in this study (Otsu, 1979; Wang et al., 2010). By using Otsu algorithm, an original gray-scale image (Fig. 2(a)) can be converted to a binary image (Fig. 2(b)), which only have two kinds of points (pure black or pure white).

In order to protect effective information (entropy) from being damaged in seismogram tracing as much as possible, we do not adopt further preprocessing methods. Furthermore, all the following processing steps are based on binary images.

2.5.2. Curves tracing

The tracing process is complicated due to different steepness of the curves in the seismograms. In our study, different tracing algorithms are used to process different types of waveforms. A technological flow chart presented in Fig. 3 shows how to trace one seismogram curve. Firstly, each curve is judged as complex type or smooth type. Secondly, all the curves of smooth type are erased from the images. Thirdly, each complex curve is traced out. Lastly, when the tracing result of one curve is not so satisfactory, a back-roll point needs to be picked out. Starting from the back-roll points, a series of key points is selected manually or automatically within a smaller range. When the tracing precision meets requirement, automatic tracing comes again.

2.5.2.1. Judging type of curves. For each curve in seismograms, we should get its start points and end points respectively, and calculate the amount of black points on one straight line including the start point and the end point. After comparing the number of black points with a default value, if the comparison result is more than 0, we decide this curve is smooth type, otherwise complex type. Finally, we trace this curve from the start point and use different methods based on its type, and record all the key points on it. In the end, the key points are stored in a database.

2.5.2.2. Tracing smooth curves. For each curve judged as smooth type on one seismogram, we take its start point $C(x_0,y_0)$ as its first key point, and the whole searching direction is horizontal right. The step length of horizontal searching is $step_x$ while the width of vertical searching is $step_y$. The $step_y$ equals to the average width of adjacent curves and can be adjusted dynamically. A sketch map in Fig. 4 shows how to trace smooth waveforms. The detailed algorithm will be described as following.

(1) Horizontal direction

The abscissa of each key point can be obtained by Eq. (1).

$$x_n = x_{n-1} + step_x \quad (1)$$

(2) Vertical direction

If $C_n(x_n, y_n)$ is a known key point, we start to judge whether point $C_n'(x_n + step_x, y_n)$ is black. If point $C_n'(x_n + step_x, y_n)$ is black, we search new black pixels nearby point $C_n'(x_n + step_x, y_n)$ and within the range $[y_n - step_y, y_n + step_y]$. When such a point of discoloration emerges, two discolor points need to be selected: $C_L(x_n + step_x, y_{nL})$ and $C_H(x_n + step_x, y_{nH})$. The value of y coordinate of the next key point is obtained by Eq. (2). If $C_n'(x_n + step_x, y_n)$ is white, we turn to trace one complex curve.