Computers & Geosciences 63 (2014) 88–95

Contents lists available at ScienceDirect

Computers & Geosciences

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

An interactive program on digitizing historical seismograms

Yihe Xu [a,b,*], Tao Xu [a]

[a] State Key Laboratory of Lithospheric Evolution, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 8 May 2013
Received in revised form
6 October 2013
Accepted 4 November 2013
Available online 16 November 2013

Keywords:
Digitization
Inversion problem
Matlab® GUI
Priori information

ABSTRACT

Retrieving information from analog seismograms is of great importance since they are considered as the unique sources that provide quantitative information of historical earthquakes. We present an algorithm for automatic digitization of the seismograms as an inversion problem that forms an interactive program using Matlab® GUI. The program integrates automatic digitization with manual digitization and users can easily switch between the two modalities and carry out different combinations for the optimal results. Several examples about applying the interactive program are given to illustrate the merits of the method.

1. Introduction

Analog seismograms are barely used nowadays, but the importance of them cannot be denied. It has only been around 70 years since the digital seismological data was recorded in the 1960s as well as the advent of WWSSN, while analog records of teleseismic data can be dated back to 1889 (Stein et al., 1988).If quantitative information could be extracted from these records, seismic events recorded during more than 70 years could be used by Geophysicists (Cadek, 1987; Samardjieva et al., 1998; Baskoutas et al., 2000; Bungum et al., 2003; Lee and Benson, 2008; Mezcua et al., 2013). New understandings may be obtained if the seismicity history is considered, because these seismic data have a longer period of time and are crucial for seismological research (Kanamori, 1988).

Digitizing analog seismograms is essential for employing historical recordings but the digitization quality directly affects the validity of final results and interpretation of the historical earthquakes. Adams and Allen (1961) used a device that is originally applied to digitize weather maps for data processing. Bogert (1961) employed equipment that processes speech and visual data for digital versions of seismograms. With the development of computers' technology and scanning and image processing techniques, digitization based on raster images scanned from photographic records was presented by Trifunac et al. (1999). Teves-Costa et al. (1999) demonstrated use of CADcore to vectorize analog seismograms and then converted the images into unevenly spaced (*x*, *y*) pairs by AUTOCAD so as to obtain the digitized seismograms. Pintore et al. (2005) developed a vectorization of historical seismograms, named *Teseo*, which bases on neural networks and Bezier fitting algorithm and provides both manual and automatic trace vectorizations of seismograms,

Although several attempts have been made for more than 70 years, and great improvements have been achieved, results from the automatic digitizing programs still have defects in some particular situations (Trifunac et al., 1973, 1999), which need subsequent corrections. It may be possible to remove these defects by purely automatic digitalization, but it needs extreme works and results in lengthy codes. The automatic tracing is faster, but its result may not be able to reach the desired quality. Manual digitization may give the best resolution, but it is time-consuming. The optimal way to digitalize analog seismograms is combination of manual and automatic tracing techniques. In this paper, we present a computer program that achieves the combination and offers a friendly graphical user interface, which allows user easily switch between the manual and automatic digitization modalities.

In the following sections, we first talk about the data we used in testing our programs, then give a general principle of automatic digitization. We described the digitization of seismograms as an inversion problem, in which the raster image acquired from scanning the historical seismograms are considered as the "observed data", and the digital seismological data saved in a sequence of amplitude data on evenly spaced time, as the "model parameters". Finally, we draw some conclusions from our implementations of the computer program and actual digitization results.

2. Data

The first step of historical seismograms digitization is scanning the analog seismograms (Trifunac et al., 1999). The output of

* Corresponding author at: State Key Laboratory of Lithospheric Evolution, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China.
E-mail addresses: xuyihe@mail.iggcas.ac.cn, blazing216@163.com (Y. Xu).

0098-3004/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2013.11.001

scanners is raster images, which are made up of two-dimensional arrays of pixels. Each pixel could be represented by either one scaled number of grayscale of pixel, or a color triplet with red, green and blue. The historical seismograms, whether it is on the smoked paper or the photographic record, are often given in grayscale colors with which our algorithm starts. Any other kinds of images could be converted into the grayscale images in terms of the values of red color. The resolution of the scanned image should be high enough to retain information sufficient for the later processing of digitization. In this study, we focused on the images that have 300 dpi and higher, since they may hold sufficient details for refinement of digitization results.

The historical seismograms are recorded by mechanical instruments, and result in images that have several characteristics, some of which increase the complexity of the problem, e.g. curvature distortion. The algorithm we presented here is based on the assumption that amplitude of seismic traces has a bijective relationship with the ordinate values, which corresponds to time. This assumption fails when dealing with the seismograms on smoked paper due to the curvature distortion (Batlló et al., 2008; Batlló, 2010; Schlupp and Cisternas, 2007). Therefore, the seismograms on the smoked paper must be particularly processed, e.g. migrating the pixels to the corrected position to reduce curvature distortion. The amount of correction of a pixel is a function of the distance between the pixel and the base line of the trace. The function might be obtained from analysis of the structure of the seismometer, or the curvature of the seismogram. We assume that the correction is already completed so that the seismograms satisfy the “bijective” assumption. We gave two examples to show capability of our program. One is a seismogram (Fig. 7) on the photographic record obtained from the paper of Pintore et al. (2005), another comes from rasterization of vector image of wide angle reflection/refraction data (Fig. 6). Although it is not historical seismograms, it has similar characteristics to the historical ones (those characteristics are discussed in the Section 3). The two examples may represent two kinds of seismograms which can be processed by our computer program. It should be mentioned that some modern seismograms may also need re-digitization to recover partial miss in the original data. Our program was originally developed to digitize wide-angle seismic reflection/refraction data.

3. Digitization

3.1. Principle

Considering the digitizing process of seismograms as a black box shown in Fig. 1, the input of the black box is a grayscale raster image and the output is seismic data. Since we need computers to do the work for us, we should offer a mathematical explanation of digitization to the computer. A grayscale raster image is merely a two-dimensional matrix of which each element has the value ranging from 0 (pure black) to 255 (pure white), and seismic traces are either parameterized lines (Pintore et al., 2005) or sequences of time–amplitude pairs (Teves-Costa et al., 1999). Modern seismological recording techniques are commonly designed for digital seismograms that contain amplitude data on evenly spaced time. The sequences of time–amplitude pairs become the format of the digital seismograms and provide possibility for further numerical process and analysis. Hence, the digitization of analog seismograms may be simply considered as a procedure that constructs a sequence or sequences of doublets from a matrix (image). If we take the matrix as “observed data” and the sequences of doublets as “model parameters”, the digitalization becomes an inversion problem.

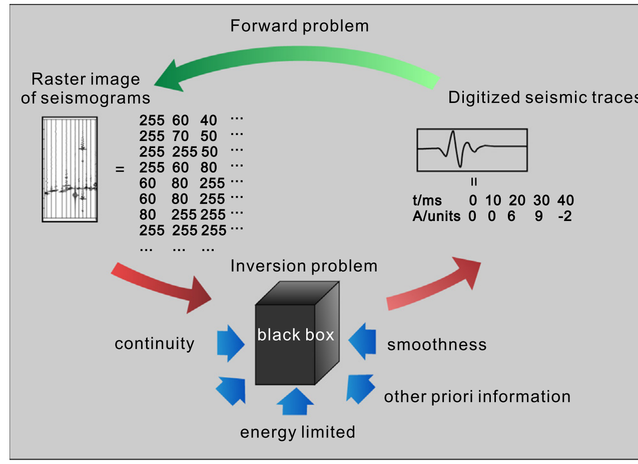

Fig. 1. The inversion problem arising in the digitization of historical seismograms. The historical seismograms can be represented by a matrix, whose elements are the grayscale value of each pixel in the scanned seismograms. The digitized seismic traces are essentially sequences of time–amplitude pairs, which contain the amplitude information in each time point. The black box can refer to any method which can retrieve seismological information from the matrix, and these methods should consider the constraints arising from the priori information.

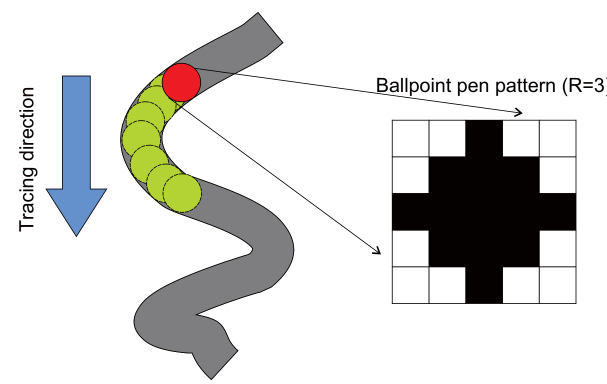

Fig. 2. Illustration of ballpoint pen assumption. The red and green balls represent ballpoint pen patterns. The red ball is the starting point, and the green ones are automatic digitization results. An example of pixel representation of ballpoint pattern is shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The relationship between the model parameters and the observed data is often defined by a mathematical operator. Mathematically, the operators must be uniquely determined by the input matrix and the output of time–amplitude pairs. However, it is difficult to give a general and explicit formula of the operator because of the requirement of detailed knowledge of the instrument structure that the authors do not have. Alternatively, we choose a simpler way that regards the forward problem of converting the seismic data to analog seismograms as moving a ballpoint pen along the traces on digital seismograms, as shown in Fig. 2. The diameter and shape of the ballpoint pen may vary as it moves. In addition, abrupt shift should be involved in case of time marks. Let $\mathbf{m}^k$ be the kth seismic traces and $\mathbf{B}$ be the ballpoint pattern, given by

$$\mathbf{m}^k = \sum_{p=1}^{Np} \delta_{im^k_{p1}} \delta_{im^k_{p2}}, \tag{1}$$

and

$$B = [b_{ij}],\ b_{ij} = \begin{cases} 1, & if \quad i^2 + j^2 \le R^2 \\ 0, & if \quad i^2 + j^2 > R^2 \end{cases}, \quad i,j = 0, \pm 1, \pm 2, \cdots, \tag{2}$$

respectively. m^k_{p1} and m^k_{p2} are both integers, and denote the abscissa and the ordinate of the pth point in the kth trace. The raster