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.

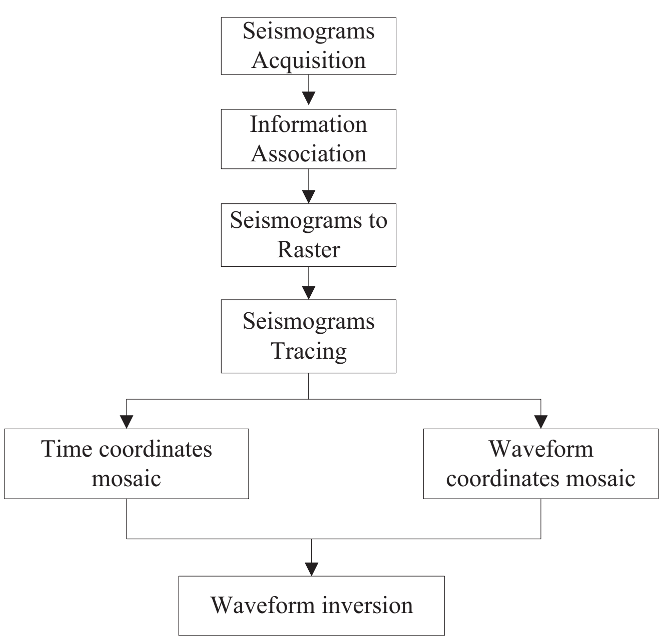

Fig. 1. An overall scheme to digitize analog seismograms.

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).

$$x_n = x_{n-1} + step_x \tag{1}$$

(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.

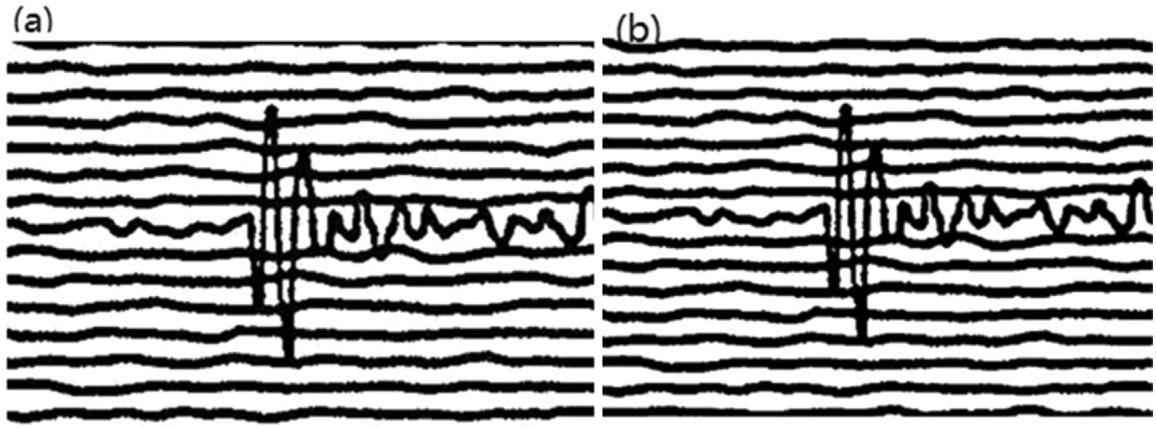

Fig. 2. (a) An original analog seismogram; (b) the processing result of Fig. 1(a) after binarization.

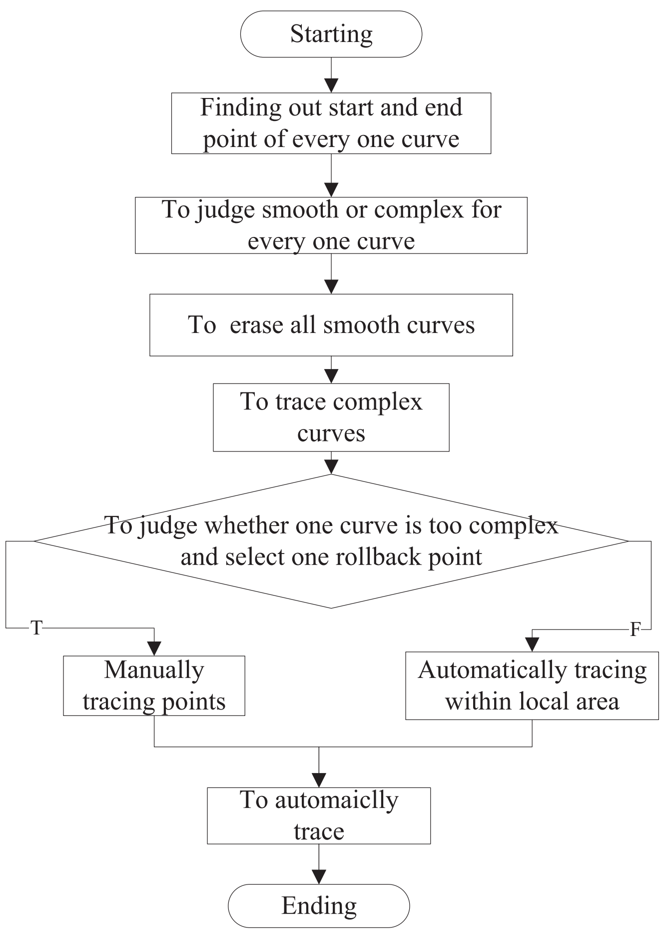

Fig. 3. A technological flow chart to trace seismogram curves.

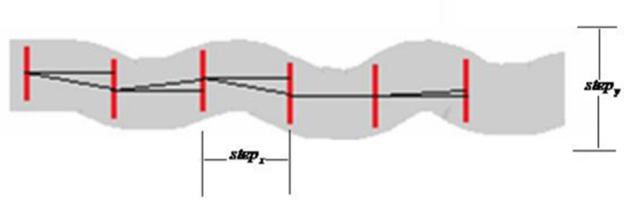

Fig. 4. A sketch map for tracing algorithm of smooth waveforms.

$$y_{n+1} = \frac{y_{nH} + y_{nL}}{2} \tag{2}$$

The Point $C_{n+1}(x_{n+1}, y_{n+1})$ is obtained based on Eq. (3)

$$\begin{cases} x_{n+1} = x_n + step_x = x_0 + (n+1)*step_x \\ y_{n+1} = \frac{y_{nH} + y_{nL}}{2} \end{cases} \tag{3}$$

An original analog seismogram with smooth waveforms is presented in Fig. 5(a). Each curve in Fig. 5(a) can be traced out by using the method of tracing smooth curves, whose result is rather satisfactory as shown in Fig. 5(b).

To avoid the interference in the following tracing of complex curves, the smooth curves traced above are wiped from original images. An original analog seismogram containing both smooth curves and complex curves is presented in Fig. 6. The smooth curves in Fig. 7 are traced out, which are marked by red dots. When all the traced smooth curves are wiped, the final result presented in Fig. 8 shows the interference level can be reduced in the tracing of complex curves.

2.5.2.3. Tracing complex curves. In the previous research, many different ways have been tried to trace complex waveforms, including tracking algorithm based on sparse pixel (Chen, 2007), tracing method with anti-noise analysis (Hang, 2000) and homotopic method following solution curve (Zhou et al., 1998), searching negative curvature points (Cao and Tang, 2007), and so on. Unfortunately, all the results seem to be not satisfying enough. After careful analysis, it is found that one complex waveform may include some smooth segments, steep wave crests, irregular cross sections, smudges, and breakpoints, etc. therefore, it is extremely difficult to find a way to deal with all the interference factors in the tracing of complex waveforms. Further analysis leads to the discovery of the phenomenon that the energy of seismic waves increases or decreases periodically and there is a point of equilibrium in between. If we can shift an observation point (one starting position for searching the next key point) slightly farther from the next key point (usually fixed on the point of wave energy equilibrium), errors and noises can be isolated in a very small range. As a result, errors do not accumulate and noises do not influence each other. Based on the above analysis, eventually, a tracing algorithm is proposed, which is named Color Scene Filed Method (CSFM). A sketch map of tracing complex curves based on CSFM is presented in Fig. 9, showing that there are six main steps, which will be described as the following.

(1) Suppose $C_n(x_n, y_n)$ is a known key point, we can judge

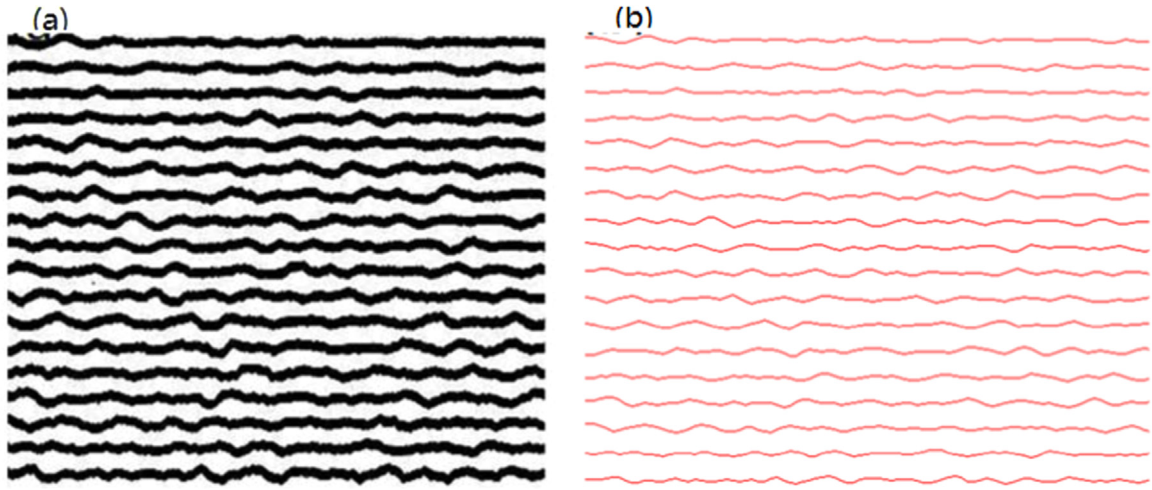

Fig. 5. (a) An original seismogram with smooth curves; (b) the tracing result of smooth curves in Fig. 5(a).

Fig. 6. An original seismogram image with smooth curves.

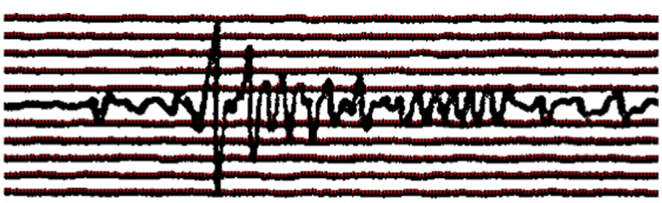

Fig. 7. A tracing result of smooth curves of Fig. 6, where smooth curves are denoted by red points.

Fig. 8. A tracing result of Fig. 6, where smooth curves are wiped from the original drawing.

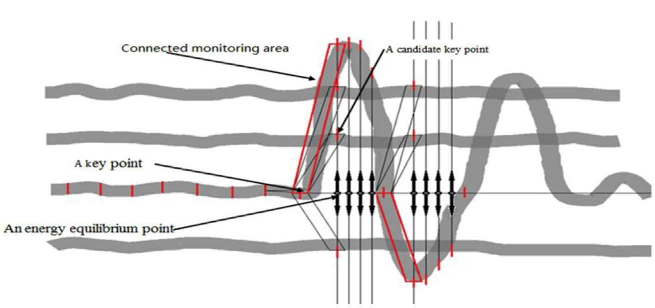

Fig. 9. A sketch map of tracing complicated waveforms based on Color scene filed method.

whether Point $C'_n(x_n + step_x, y_n)$ is black, if it is black, turn to smooth curve tracing and being wiped discussed above.

(2) If $C'_n(x_n + step_x, y_n)$is white, we search black pixels nearby Point $C''_n(x_n + step_x, y_0)$ within the range $[y_0 - step, y_0 + step]$, and Point $C''_n(x_n + step_x, y_0)$ is usually put on one wave energy equilibrium position. Then y values of all the low boundary points of connected black areas are saved into *set* $a_{down}[i](i = 0, 1, \ldots, m)$, and y values of all the upper boundary points of connected black areas are saved into *set* $a_{up}[i](i = m + 1, \ldots)$, so we get a points set $C'_{n+1}(x_{n+1}, a[i])$ $(i = 0, 1, m, m + 1 \ldots)$.

(3) We can establish many groups of four points: $(x_n - \Delta, y_n)$, $(x_n + \Delta, y_n)$, $(x_{n+1} - \Delta, a[i])$, and $(x_{n+1} + \Delta, a[i])$ $(i = 0, 1, \ldots)$ where Δ is an dynamically empirical value. For each candidate key point $C'_{n+1}(x_{n+1}, a[i])$, we can get one parallelogram with the four points as vertexes, whose area value is expressed as Eq. (4).

$$S_i = \Delta^* |a[i] - y_n| \tag{4}$$

(4) Then we calculate out the number of black points denoted by n_i in each one parallelogram above. The ratio of n_i to the area of the corresponding parallelogram denoted by η_i can be computed by Eq. (5).

$$\eta_i = \frac{n_i}{S_i} \tag{5}$$

The larger the value of η_i to one candidate point $C'_{n+1}(x_{n+1}, a[i])$ is, the better the connectivity between the candidate key point and the previous key point $C_n(x_n, y_n)$ becomes. So the point $C'_{n+1}(x_{n+1}, a[i])$ owning the maximum value of η_i is fixed as the next key point.

(5) If $C_{n+1}(x_{n+1}, y_{n+1})$ is a new key point searched by the above steps, we start to judge whether point $C'_{n+1}(x_{n+1} + l, y_{n+1})$ $(l < \Delta)$ is black. If point $C'_{n+1}(x_{n+1} + l, y_{n+1})$ is white, turn to smooth curves searching which is discussed above. If point $C'_{n+1}(x_{n+1} + l, y_{n+1})$ is black, point $C''_{n+1}(x_{n+1} + l, y_0)$ is taken as one new energy equilibrium position. And starting to search beginning from which by steps (2–4), lastly we get the next key point $C_{n+2}(x_{n+2}, y_{n+2})$.

(6) Repeat steps(1–5), until all the key points on the whole curve are detected.

2.6. Examples of tracing results

The first example is to digitize an analog seismogram containing both smooth curves and a small amplitude curve (Fig. 10

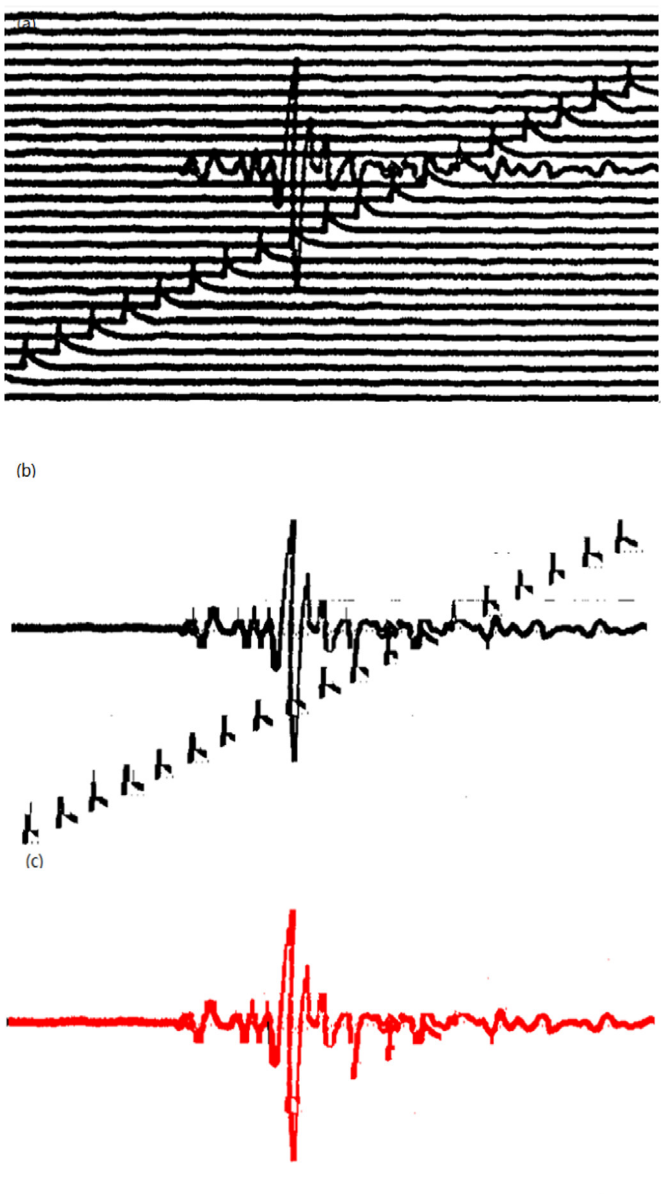

Fig. 10. (a) An analog seismogram with smooth and complicated curves; (b) smooth curves are wiped from the original drawing; (c) a complicated curve are tracing out based on (b).

(a)). First, all the smooth curves are wiped from the original image, and the complex curve are left with the time points and noise points (Fig. 10(b)). Next, we trace out the complex curve (Fig. 10(c)) based on CSFM. The tracing results show that the basic topological structure of the original complex wave has been well preserved.

The second example is to digitize an analog seismogram containing a more complex curve (Fig. 11(a)), in which there is much loss of correspondence between the abscissa and time, because the trace at its maximum amplitude is somewhat ahead of the zero crossing at the same time. First, all the smooth curves are wiped from the original image and the complex curve are left (Fig. 11(b)). Second, the complex curve (Fig. 11(b)) is traced out based on CSFM, and the basic topological structure of the original complex wave is also well preserved. Correction between the abscissa and time is worthy of further study, and will be researched in the future.

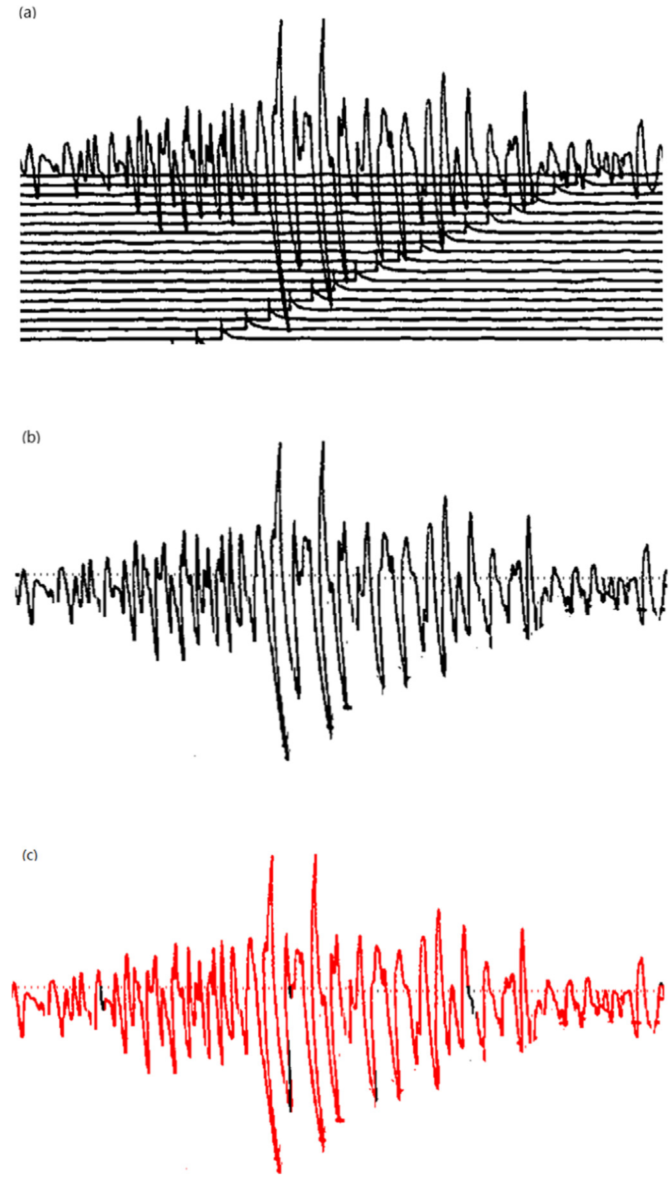

Fig. 11. (a) An analog seismogram with smooth and complicated curves; (b) smooth curves are wiped from the original drawing; (c) a complicated curve are tracing out based (b).

2.7. *Mosaic and inversion*

2.7.1. *Coordinate mosaic*

Coordinate mosaic method presented in our previous work (Wang et al., 2014) will be described in the following.

After vectorization of one paper seismogram, all the points on the seismogram are saved into many arrays a[i] (i=1, 2, …, n) with two dimensions, where the points on the one curve are stored in one array.

A sketch map of coordinate mosaic presented in Fig. 12 shows that the last point on a[0] curve coincides with the first point on a

Fig. 12. A sketch map of coordinate mosaic.

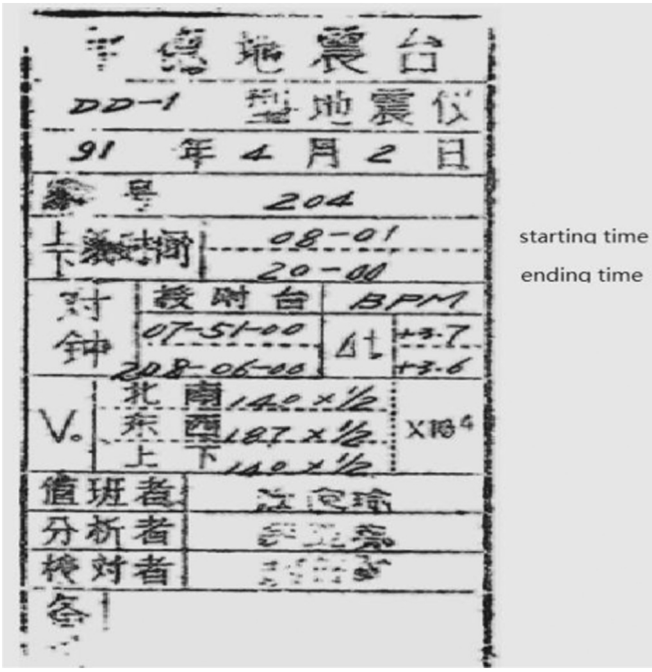

Fig. 13. An additional information table for an analog seismogram.

[1] curve, and the last point on a[1] curve coincides with the first point on a[2] curve, and so on. So we can get the (Eqs. (6)–7).

$$\therefore a[i].x' = a[i].x - \sum_{n=0}^{i-1} \Delta x_n \tag{6}$$

Here Δx_n is the x coordinate difference between the first point on the nth curve and the last point on n-1th curve.

$$\therefore a[i].y' = a[i].y - \sum_{n=0}^{i-1} \Delta y_n \tag{7}$$

Here Δy_n is the y coordinate difference between the first point on the nth curve and the last point on $n-1$th curve.

All points on one seismogram can be normalized onto one curve with continuous coordinates by (Eqs. (6)–7).

2.7.2. Time mosaic

For an analog seismogram, start and end time is recorded in a table as shown by Fig. 13, using which, we hope to map all the time values for every one points on paper seismograms. How to compute time values of corresponding coordinate points on normalized curves based on these hand-marked time points is another key problem.

For one hand-marked time point $b[j]$ (j=0, 1… m), we can extract one waveform wave $a[i]$ (i=0, 1… n) which has a minimum distance to the hand-marked time point just as described by Eq. (8).

For each point on one normalized curve, we can get its time value by Eq. (8).

$$Time_i = Time_{start} + (x_i - x_{start}) \times \frac{Time_{end} - Time_{start}}{x_{end} - x_{start}} \tag{8}$$

In Eq. (8), $Time_i$ is the recording time of the ith point on one normalized curve, $Time_{\mathrm{start}}$ and $Time_{\mathrm{end}}$ are the start time and end time of one whole recording respectively, and x_{start} and x_{end} are the x coordinate value of the first point and the last point on one normalized curve.

Fig. 14. Seismogram inversion.

2.7.3. Inversion

Waveform inversion algorithm method has also been discussed (Wang et al., 2014), here we give the newest optimization technology.

Firstly, we do not need to get all the points of one normalized curve from database in one inversion, and the point-number can be decided and loaded dynamically as shown in Fig. 13, which minimizes computer usage of RAM.

In addition, the inversion interface becomes more user-friendly by providing the ability to zoom in and out as shown in Fig. 14.

3. Conclusion

This paper mainly focuses on how to digitize analog seismograms and a new curve-tracing method is provided. No classical methods have been used to digitize analog seismograms. However, based on the final tracing results of smooth and complex curves as shown above, the topology structures of tracing results are found to match the corresponding analog seismograms. Moreover, an effective mosaic algorithm is proposed. In the end, what comes out of the research is that a software prototype has been designed and developed. The altogether 10G analog seismograms, which come from Chengde Seismic station in 1991, have been digitalized by using the software.

The project team has been devoted to pursuing open source in the past. The task of our upcoming study will be on how to improve the algorithm and refine the software. It is hoped that two modules will be developed in the near future, one of which intends to correct the digital results of analog seismograms, and another module intends to identify and extract different types of waveforms automatically.

Acknowledgment

We are grateful to professor Jilin Feng from the Institute of Disaster Prevention for his helpful suggestion to improve the quality of this work. We particularly thank the Supported by National Natural Science Foundation of China (Researches on key algorithms and software development in analog seismogram records vectorization, 41504037).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http: //dx.doi.org/10.1016/j.cageo.2016.05.004.

References

Bogert, B.P., 1961. Seismicdata collection, reduction, and digitization. Bull. Seism. Soc. Am. 51 (4), 515–525.

Bungum, H., Lindholm, C., et al., 2003. Long-period ground-motions for large European earthquakes, 1905–1992, and comparisons with stochastic predictions. J. Seism. 7 (3), 377–396.

Cao, D.H., Tang, J.K., 2007. Algorithm for fast extracting human limb contours using searching NCM points. J. Huazhong Univ. Sci. Technol. (Nat. Sci. Ed.) 35 (5), 16–18.

Chen, B.F., 2007. Research and Implementation of Log Curve Vectorization Methods and Rebuilding Technique. Harbin Institute of Technology, China.

Hang, Z., 2000. Automatic Recognition and Tracing of Curves on Well Log Graphs. University of Sichuan, China.

Harvard station, 2015. ⟨http://www.seismology.harvard.edu/resources.html#HRV⟩.

Otsu, N., 1979. A Threshold selection method from gray-level histograms. IEEE Trans. Syst. Man. Cybern. 9 (1), 62–66.

Pintore, S., Quintiliani, M., et al., 2005. Teseo: a vectoriser of historica seismograms. Comput. Geosci. 31 (10), 1277–1285.

Teves-Costa, P., Borges, J., Rio, et al., 1999. Source parameters of old earthquakes: semi-automatic digitization of analog records and seismic moment assessment. Nat. Hazards 19 (2–3), 205–220.

Trifunac, M.D., Lee, V., et al., 1999. Common problems in automatic digitization of strong motion accelerograms. Soil. Dyn. Earthq. Eng. 18 (7), 519–530.

Wang, M.F., Zou, X.P., et al., 2010. Recognition of atomically-resolved STM images of graphite. J. Comput. Theor. Nanosci. 7, 404–407.

Wang, M.F., Jiang, Q.G., Feng, J.L., et al., 2014. A new waveform mosaic algorithm in the vectorization of paper seismograms. Sens. Transducers 182 (11), 203–206.

Xu, J.S., Xu, ks, et al., 2008. Saving and data sharing of historical seismogram in Beijing National Earth Observatory. Seism. Geomagn. Obs. Res. 29 (3), 100–104.

Xu, Y.H., Xu, T., 2014. An interactive program on digitizing historical seismograms. Comput. Geosci. 63, 88–95.

Zhou, Z., Zhu, X.P., et al., 1998. A new homotopic algorithm following solution curve. J. Northwest. Polytech. Univ. 16 (2), 246–250.