Adaptive real-time spectrum data compression and recovery method

mingsheng zhou[1], minging kong[1], Zheng Pei[1], Junkai Xiong[1], Yuling Tang[1], and Binbin Deng[1]

[1]Xihua University

November 23, 2022

Abstract

This paper presents an adaptive real-time spectrum data compression and recovery method for radio monitoring, its purpose is to reduce the data storage space and the network bandwidth occupied by transmission without affecting the subsequent analysis and application. Considering the similarity between spectrum data, we use correlation coefficients and bitmap similarity to measure them, and then replace all the original spectrum with a small amount of typical spectrum to achieve the purpose of compressing the original spectrum for storage and compression. The experimental conclusions show that the method can automatically adapt to various radio frequency bands and achieve better compression effects.

Adaptive real-time spectrum data compression and recovery method

Mingsheng Zhou[1], Mingming Kong[1*], Zheng Pei[1],

Junkai Xiong[1], Yuling Tang[2], Binbin Deng[1]

[1]Center for Radio Administration & Technology Development, Xihua University, Chengdu 610039, Sichuan, China
[2]Sichuan Radio Monitoring Station, Chengdu, 610016 China

Corresponding author: Mingming Kong (kongming000@126.com)

Key Points:

- The algorithm is used to judge the similarity of the spectrum, and the similarity threshold is adaptive to any radio frequency band.
- The Pearson similarity coefficient judges the similarity of the signal envelope.
- The bitmap similarity coefficient judges the degree of signal overlap.

Abstract

This paper presents an adaptive real-time spectrum data compression and recovery method for radio monitoring, its purpose is to reduce the data storage space and the network bandwidth occupied by transmission without affecting the subsequent analysis and application. Considering the similarity between spectrum data, we use correlation coefficients and bitmap similarity to measure them, and then replace all the original spectrum with a small amount of typical spectrum to achieve the purpose of compressing the original spectrum for storage and compression. The experimental conclusions show that the method can automatically adapt to various radio frequency bands and achieve better compression effects.

Keywords: Spectrum data compression; Correlation coefficient; Real-time; Adaptive

1. Introduction

In the field of radio monitoring, spectrum data is the main research object and plays an important role in radio spectrum management. A lot of meaningful spectrum information can be obtained by analyzing scanning spectrum data [1-2], which has always been an important means to effectively manage spectrum resources and improve radio spectrum utilization. In practical applications, the acquisition of useful information in the radio spectrum is mainly through real-time spectrum data analysis, or data mining of past spectrum data [3]. Therefore, we need to store radio signal spectrum data for a long time, however, this leads to the need for a large amount of storage media consumption.

To efficiently store the huge amount of spectrum data, many spectrum data compression algorithms have been proposed, such as inspired by image compression methods, several improved image compression algorithms have been used to compress spectrum data [4-6]. In [7-8], spectrum data compression algorithms are based on energy detection, that is the algorithms realize the compression and storage of spectrum data by separating signal and noise. Summary, existing algorithms commonly possess the following characteristic: 1) they are relied on prior knowledge, i.e., by analyzing the pre-stored spectrum data and extracting information from the pre-stored spectrum data, these algorithms are utilized to compress and store spectrum data. 2) they have a high time complexity and recovery distortion. 3) they are used to compress and store off-line spectrum data, and there is no algorithm can be used to compress and store real-time or on-line spectrum data.

In this paper, an adaptive real-time spectrum data compression and recovery method is proposed, in which similarity among radio spectrum data is utilized. In applications, it is worth to notice that based on different similarity measurements in different frequency bands, the new spectrum data compression algorithm has low time complexity, the advantage can be used to avoid affecting the collect of spectrum data in real-time compression [9]. In addition, according to comparison experiments, the proposed spectrum data compression algorithm is more effective to handle real-time network spectrum data transmission than existing algorithms.

The rest of the paper is structured as follows: In Section 1, the similarity between radio spectrum data is introduced, including: spectrum data storage method and the process of spectrum similarity measurement, minimum similarity coefficient, correlation coefficient and bitmap similarity. In Section 2, an improved similarity measurement algorithm based on the normal distribution of noise is proposed. In Section 3, the compressed and restored spectrum data is shown.

In Section 4, the performance of the improved algorithm is verified through experiments. Section 5 concludes that the algorithm can perform real-time adaptive compression of spectrum data without serious distortion.

2. Similarity among radio spectrum data

2.1. Similarity measurement process

To achieve spectrum data compression, first, obtain a frame of original spectrum data, use the original spectrum to compare the similarity with all typical spectrums data in the typical spectrum database one by one. If the original spectrum is similar to a typical spectrum in the database, use the serial number of the typical spectrum in the database to replace the original spectrum, otherwise, compare with the next typical spectrum. If all the typical spectrums are not similar to the original spectrum, it is added to the typical spectrum database and assigned a number.

The basic idea is to judge the similarity of the signal envelope of the two frames spectrum, then judge the overlap degree of the region in the envelope, and determine whether it is similar or not according to the comprehensive results. Based on this, the judgment criteria for the similarity between two frames of spectrum data is that the following two conditions are all satisfied in turn: $R_s > R_{min}$ and $R_b > R_{min}$. That is, if the similarity coefficient R_s is greater than the minimum similarity coefficient R_{min}, then the bitmap similarity R_b is also judged, if it is greater than R_{min} too, the two frames of spectrum data are similar, otherwise it is not similar. As show in Fig.1.

The similarity coefficient R_s is obtained by calculating Pearson similarity coefficient of two frames of spectrum data, R_b is the bitmap similarity. And the minimum similarity coefficient R_{min} is obtained by empirical formula, which is obtained based on the characteristics of the spectrum data, obviously, the similarity judgment process is adaptive.

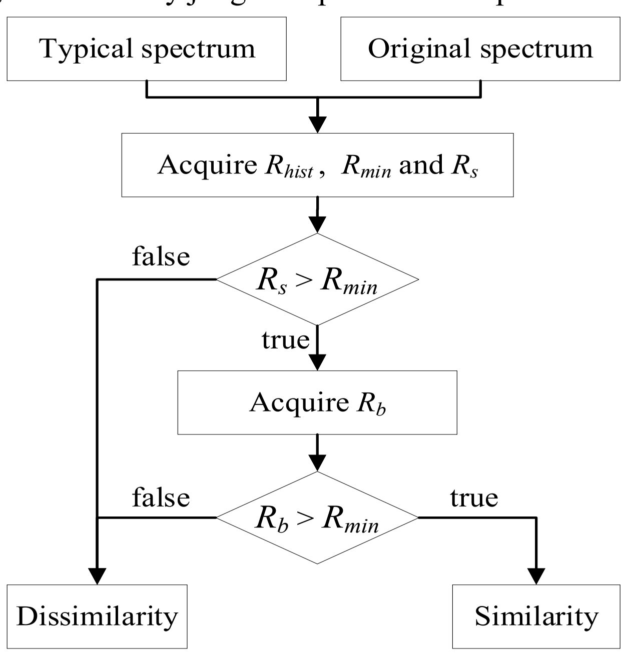


Fig.1. Similarity measurement of process 1

2.2. Correlation coefficient

Since spectrum data can be regarded as a vector, the similarity of spectrum can be measured by those methods that measuring the similarity between two vectors. Existing vector distance calculation methods, such as Euclidean distance, calculate the total difference between two vectors [10], which cannot give expression to the shape similarity of data. In this paper, the Pearson correlation coefficient is used, and it can represent whether the trend of change of two vectors is the same [11]. Therefore, when applied to the spectrum data, we can judge whether two frames of spectrum are linearly correlated, that is, whether their shapes are similar, and then make a further judgment. The formula of Pearson correlation coefficient is as follows:

$$\rho_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma_X}\right)\left(\frac{Y_i-\overline{Y}}{\sigma_Y}\right) \tag{1}$$

Where ρ_{XY} is the Pearson correlation coefficient value, X_i and Y_i are the $i-th$ of the two vectors, $\overline{X}$ and $\overline{Y}$ is the average value, σ_X and σ_Y is the standard deviation of the two vectors, respectively.

2.3. Minimum similarity coefficient

The minimum similarity coefficient is used to determine whether two frames of spectrum are similar. In actual radio monitoring, the number of signals and noise parts in different frequency bands are different. Even in the same frequency band, the frequency band occupancy of signals and noise at different times are constantly changing.

In the process of radio monitoring, more attention is paid to the signal parts of spectrum data. Therefore, during spectrum data compression, the information of the signal part should be kept as much as possible. In other words, the similarity between spectrum mainly requires that the signal parts must be similar to a great extent without paying too much attention to the noise part. So, we propose an adaptive method to obtain R_{min}, i.e., according to the data changes of different frequency bands to achieve a better compression effect. That is, R_{min} is dynamically adjusted by the proportion of noise in the whole given frequency band.

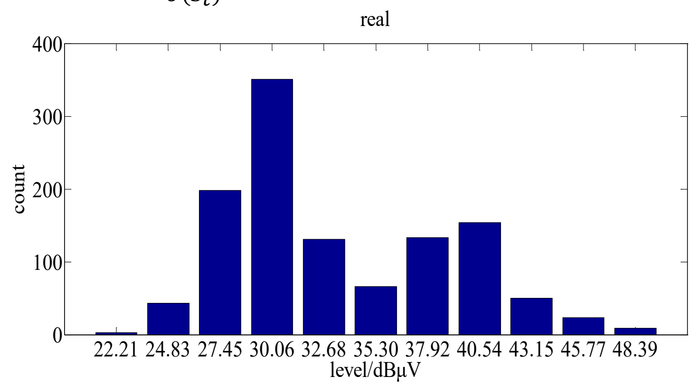
Empirically, the more the spectrum data obeys the normal distribution [12-13], the greater the proportion of noise, on the contrary, the more the number of signals. So, the proportion of noise can be judged according to the degree that it obeys the normal distribution. We translate this question into: compare whether the histogram of the spectrum data is similar to the histogram from the normal distribution [14]. If the similarity R_{hist} between the two histograms is larger, the proportion of noise is larger, and the coefficient R_{min} should be small relatively. On the contrary, if R_{hist} is smaller, then indicate the number of signals is more, here, in order to retain as much signal information as possible in the compression process, the coefficient R_{min} is required to be large relatively. Therefore, we use the following empirical formula to calculate R_{min}.

$$R_{min} = 1 - (0.3 + 0.7 \times R_{hist}) \tag{2}$$

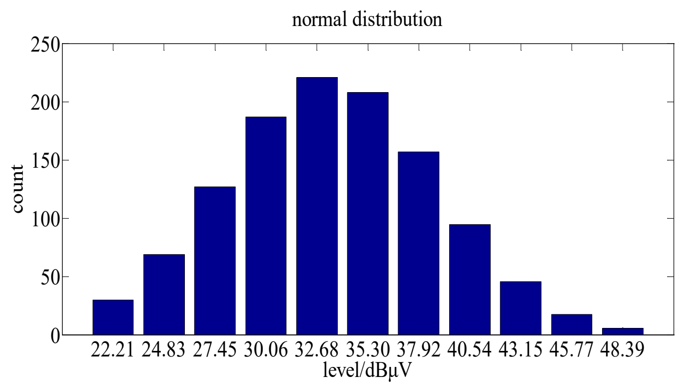
R_{min} and R_{hist} is negative correlation, 0.7 represents the weight in the negative correlation, 0.3 for limiting the maximum range of R_{min} will not exceed 0.7.

Formally, for a frame of original spectrum data $S_t = \{s_1, s_2, \cdots, s_i, \cdots, s_n\}$, where s_i is the level value of this frame spectrum. Firstly, divide the range $[S_{min}, S_{max}]$ of level values into several subintervals $G = \frac{S_{max}-S_{min}}{g}$, in which, $S_{max} = max(S_t)$, $S_{min} = min(S_t)$, g is the number of intervals, which is taken as 11 in this paper.

Second, for every s_i of one frame of spectrum, statistic the number of spectrum data points placed in each subinterval $H_t = \{C_1, C_2, \cdots, C_i, \cdots, C_{11}\}$,then get the histogram of H_t, as shown in Fig2(a). Meanwhile, calculate the median value of each subinterval $M_t = \{M_1, M_2, \cdots, M_i, \cdots, M_{11}\}$, respectively. Thirdly, calculate the mean $S_{Avg} = Avg(S_t)$ and variance $\sigma(S_t)$ of S_t, and the normal distribution of the current spectrum is obtained according to the $Avg(S_t)$ and $\sigma(S_t)$: $F(x) = \frac{x-Avg(S_t)}{\sigma(S_t)}$.



(a) Real histogram



(b) Normal distribution histogram

Fig.2. Real histogram and Normal distribution histogram of original spectrum data.

Let $X = \frac{n}{\sum_{i=1}^{11} P_i}$ be the total number for M_t in each subinterval of real histograms corresponding to the normal distribution histogram, where n is total number of S_t, $P_i = \frac{M_i-Avg(S_t)}{\sigma(t)}$, P is probability value of each subinterval, and then $V_i = P_i \times X$, as shown in Fig2(b). $V = \{V_1, V_2, \cdots, V_i, \cdots, V_{11}\}$ is the amount of data of each subinterval. Calculate the Pearson correlation coefficient between vector H_t and vector V, when the correlation coefficient $R_{hist} >$ 0.8, it is determined that the frame spectrum obeys the normal distribution. Finally, according to the correlation coefficient of histogram R_{hist}, the minimum similarity coefficient R_{min} can be obtained.

2.4. Bitmap similarity

Pearson correlation coefficient can be used to preliminary judge whether the waveforms of the spectrum are similar, but even if the waveforms are similar, there may be a large gap between the corresponding level values of the two frames of spectrum data, so it is impossible to directly assert whether the two frames of spectrum are similar.

To solve this problem, after using the Pearson correlation coefficient R_s of the two frames of spectrum data to judge that they are similar, we convert the spectrum to a binary bitmap, and continue to perform a further comparison method to determine whether the two frames of spectrum are similar. That is, by comparing Pearson correlation coefficient of two binary bitmaps.

The essence of converting spectrum data into binary bitmap is to convert 1-dimensional data S_t into 2-dimensional binary matrix B[row][col]. The construction process of 2-dimensional matrix B is: assign values to each element in the matrix and compare them with the spectrum data one by one. If it is greater than the corresponding level value, it will be set to 0, otherwise it will be set to 1. Through this operation, the area above the spectrum envelope will be filled with white and the area below will be filled with black, the build process as show in Algorithm 1. Where $row = \frac{S_{max}-S_{min}}{step}$, let $step = 0.5$, S_{max} and S_{min} is the upper limit and lower limit of the level value of a given frequency band, which obtained from multi frame spectrum data through monitoring for a period of time. And col is the number of points of S_t.

A spectrum bitmap of the broadcast frequency band is shown in Fig3, in which the values of the black area be 1 and the values of the white area be 0.

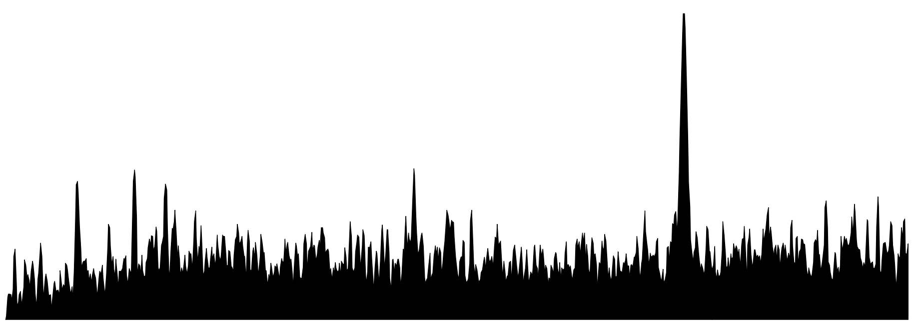


Fig.3. Spectrum bitmap in broadcast band.

Algorithm 1: Spectrum data convert to binary bitmap

Input: spectrum data array S_t

Output: 2d array of binary bitmaps B

```
for var=S_min and i=0 ; var<S_max ; var+=step and i++ do
    for j=0; j<col; j++ do
        if var > S[j] then
            B[i][j] = 0
        else
            B[i][j] = 1
        end if
    end for
end for
```

The similarity of binary bitmaps is the quotient of their intersection divide by union [15]. The formula is as follows:

$$R_b = \frac{B_{orig} \cap B_{typi}}{B_{orig} \cup B_{typi}} \tag{3}$$

Where R_b is bitmap similarity, B_{orig} and B_{typi} is the filled area with black in the original spectrum bitmap and the typical spectrum bitmap respectively. Then, $U = B_{orig} \cup B_{typi}$, is the number of elements of the union set that value be 1 in both two bitmaps, $A = B_{orig} \cap B_{typi}$, is the number of elements of the intersection set that value be 1 in two bitmaps. The calculation process is shown in Algorithm 2.

Algorithm 2: Bitmap similarity

Input: bitmap B_{orig} and B_{typi}

Output: bitmap similarity R_b

```
for i=0; i<row; i++ do
    for j=0; j <col; j++ do
        if B_orig[i][j] == 1 and B_typi[i][j] == 1 then
            A++, U++
        else if B_orig[i][j] ≠ B_typi[i][j]
            U++
        end if
    end for
end for
R_b=A/U
```

3. Optimization of Similarity Measure

Experiments show that when most of the spectrum data are noise and only a few small signals, as shown in Fig4, the compressed spectrum often has distortion, especially in these signal areas. Here, we improve the algorithm.

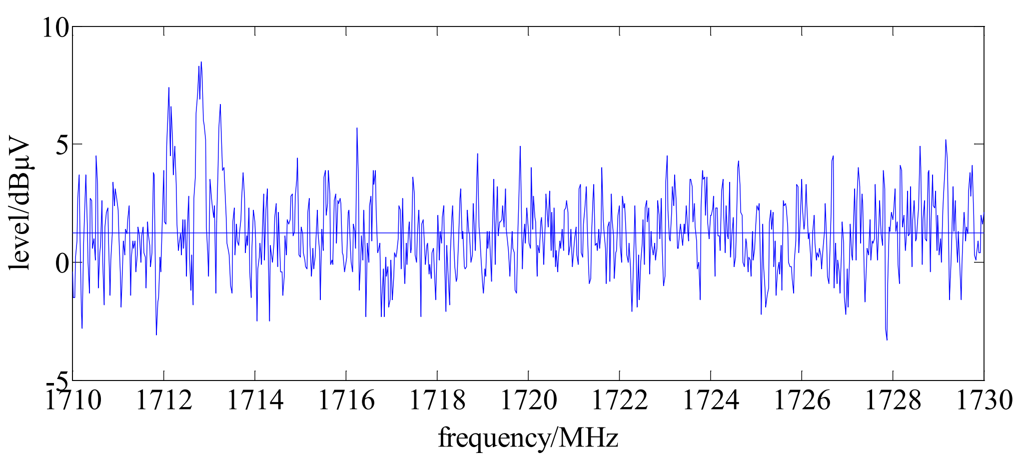


Fig.4. Spectrum of a small number of signals

Firstly, judge whether the original spectrum data obey the normal distribution, that is, judge whether R_{hist} is greater than the given threshold $P = 0.8$. If true, it indicates that the noise occupies the majority in this frame of spectrum. Then, judge whether there are signals in the frame spectrum. If the maximum level value S_{max} of the frame 5 greater than the average value, that is $S_{max} - S_{Avg} > 5$, it is determined that there are some signals [16]. Thirdly, Judge whether the bitmap similarity R_b is greater than the given threshold $R_{min} = 0.4$. If it is true, it is asserted that they are similar. It is worth noting that, here we use a skill on obtaining bitmap similarity, that is, let $S_{min} = S_{Avg}$ during bitmap transformation, so that these signals in the bitmap occupies the main part, which reduces the impact of noise on bitmap similarity.

When it does not obey the normal distribution, that is, $R_{hist} < P$, it indicates that the signals in the spectrum occupy the majority, and the spectrum is compressed according to the previous judgment process of similarity. The similarity measurement process described in Fig5.

Due to the redundancy of spectrum data, a limited number of typical spectrums always can be found to represent all original spectrum. After similarity judgment, if the original spectrum is similar to the typical spectrum, the serial number of the typical spectrum can be used to replace the original spectrum. If all the typical spectra are not similar to the original spectrum, the original spectrum can be used as a new typical spectrum. The result of similarity judgment is: either the serial number of the typical spectrum is obtained, or the new typical spectrum and its serial number are obtained. Then, the content of the stored files is the typical spectrum stored by serial and the serial list corresponding to the original spectrum.

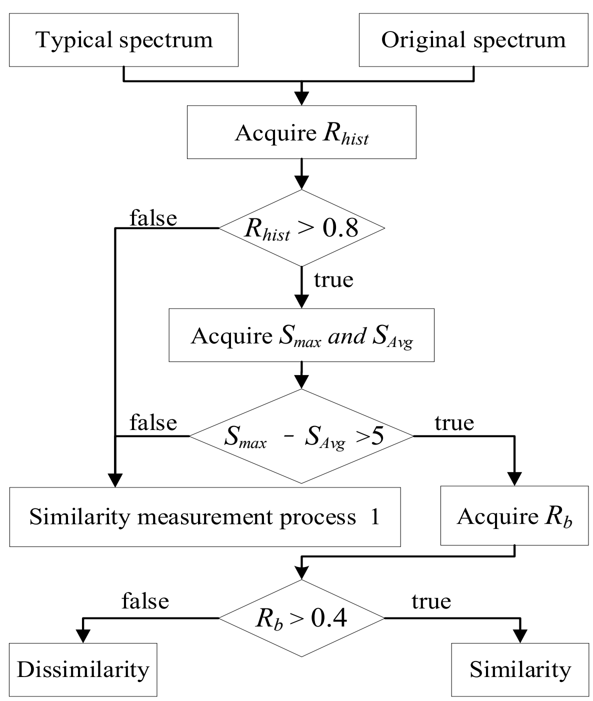


Fig.5. Similarity measurement of process 2.

4. Compression and recovery

When recovering the spectrum data, the serial number list is read from the file, and the typical spectrum data is found by the serial number and replaced in turn. Of course, this algorithm is a lossy compression, but from the perspective of radio monitoring, almost all the useful information used for frequency band occupancy, time occupancy analysis and electromagnetic environment evaluation are retained, as show in Fig6, so we can tolerate it.

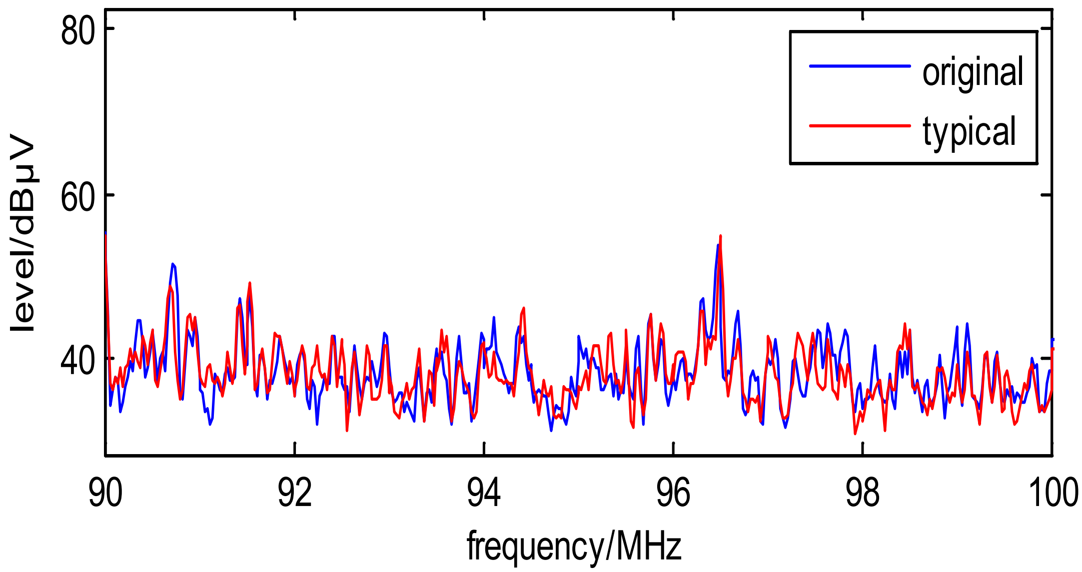


Fig. 6. The comparison of original spectrum and typical spectrum.

5. Experiment

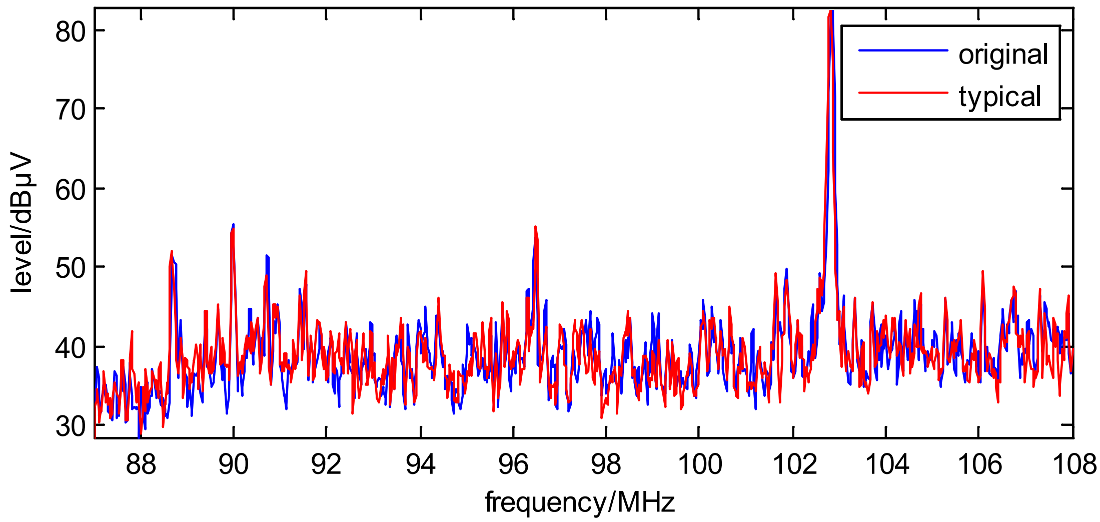
The hardware platform of this experiment is composed of EM100 receiver, HE600 antenna, and PC. After receiving the wireless spectrum data packet, EM100 receiver sends it to PC through LAN. Then, the program on PC communicates with EM100 based on SCPI protocol. After receiving the data packet, the frequency, level values and other relevant information can be obtained by parsing according to the protocol. While receiving spectrum data, each received frame spectrum can be compressed according to the algorithm.

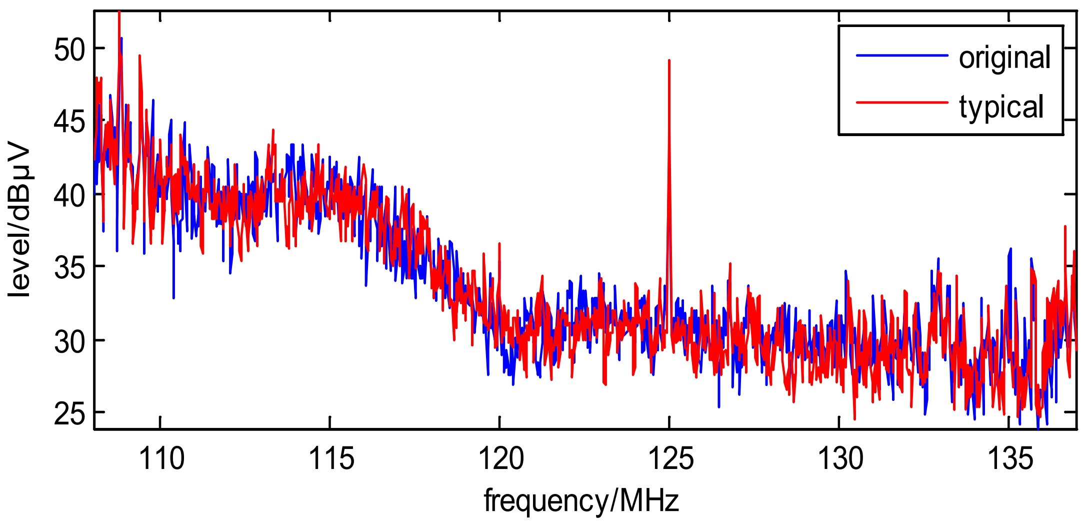
TABLE 1

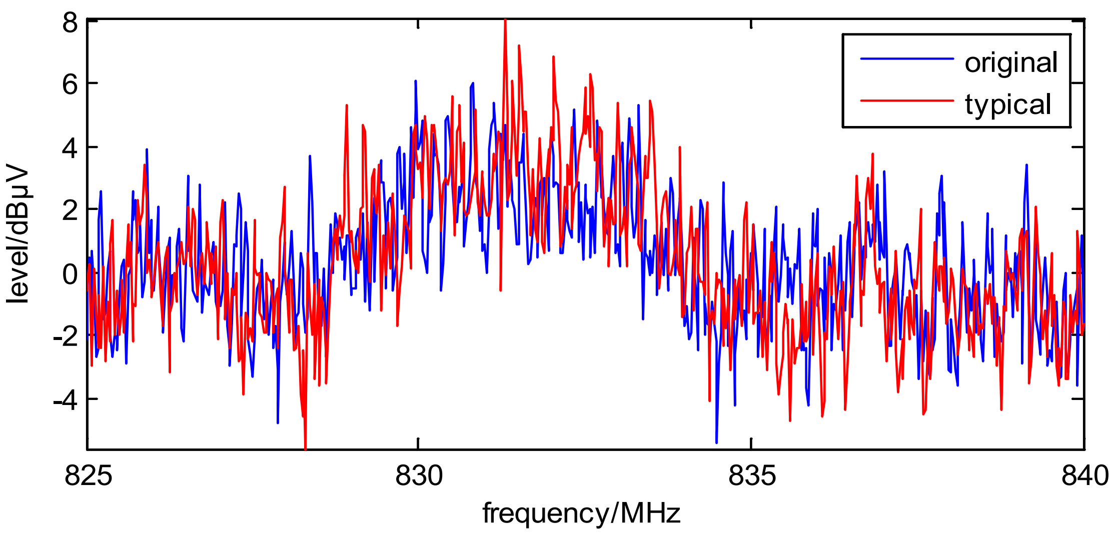
COMPRESSION RATIO AND ERROR IN DIFFERENT BANDS.

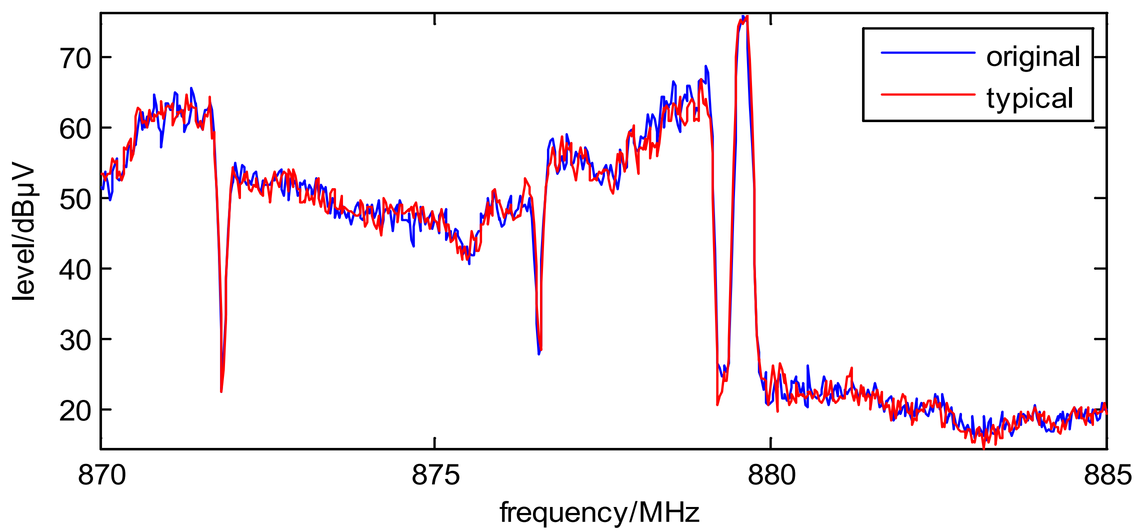
Band/MHz	***CR***	***PRD***	***SRD***
87-108	0.17	8.58	62.14
108-137	0.24	8.67	63.68
825-840	4.93	112.38	121.65
870-885	0.89	6.97	18.18
890-909	5.67	78.50	88.68
935-954	1.23	19.74	48.68
1710-1755	55.49	37.64	52.67
1900-1920	5.28	13.84	25.95
2635-2655	35.48	75.00	79.83

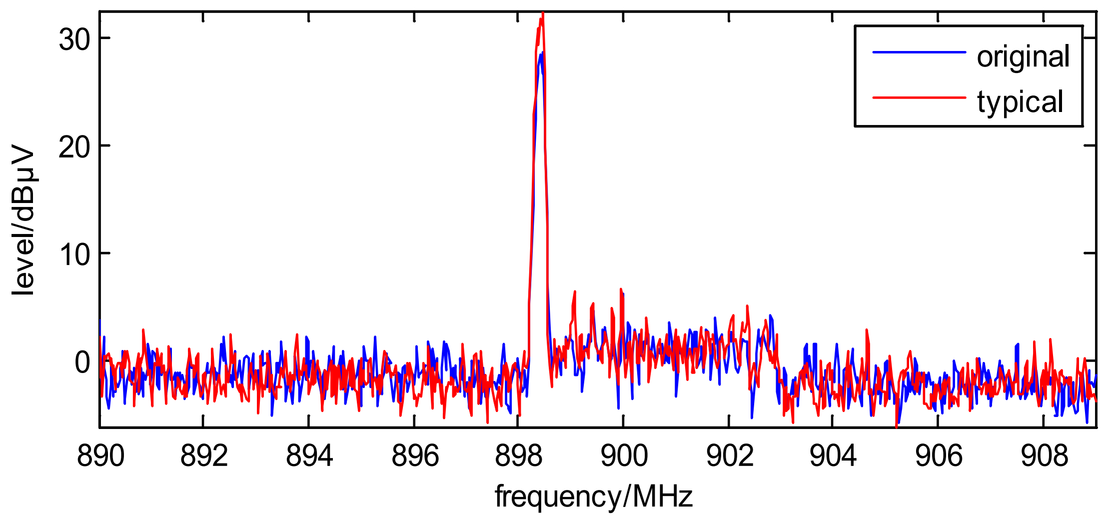
After experiments on different frequency bands, the original spectrum files and compressed files are obtained respectively. And the original spectrum and typical spectrum in each band are shown in Fig7.

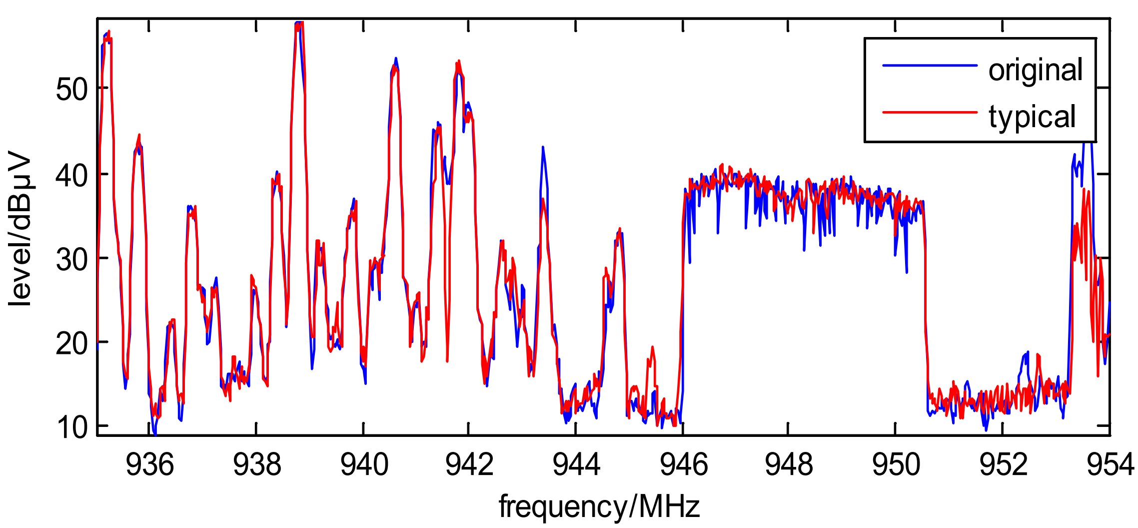

original
typical
level/dBμV
80
70
60
50
40
30
88
90
92
94
96
98
100
102
104
106
108
frequency/MHz


original
typical
level/dBμV
50
45
40
35
30
25
110
115
120
125
130
135
frequency/MHz


original
typical
level/dBμV
8
6
4
2
0
-2
-4
825
830
835
840
frequency/MHz


original
typical
level/dBµV
70
60
50
40
30
20
870
875
880
885
frequency/MHz


original
typical
level/dBµV
30
20
10
0
890
892
894
896
898
900
902
904
906
908
frequency/MHz


original
typical
level/dBµV
50
40
30
20
10
936
938
940
942
944
946
948
950
952
954
frequency/MHz

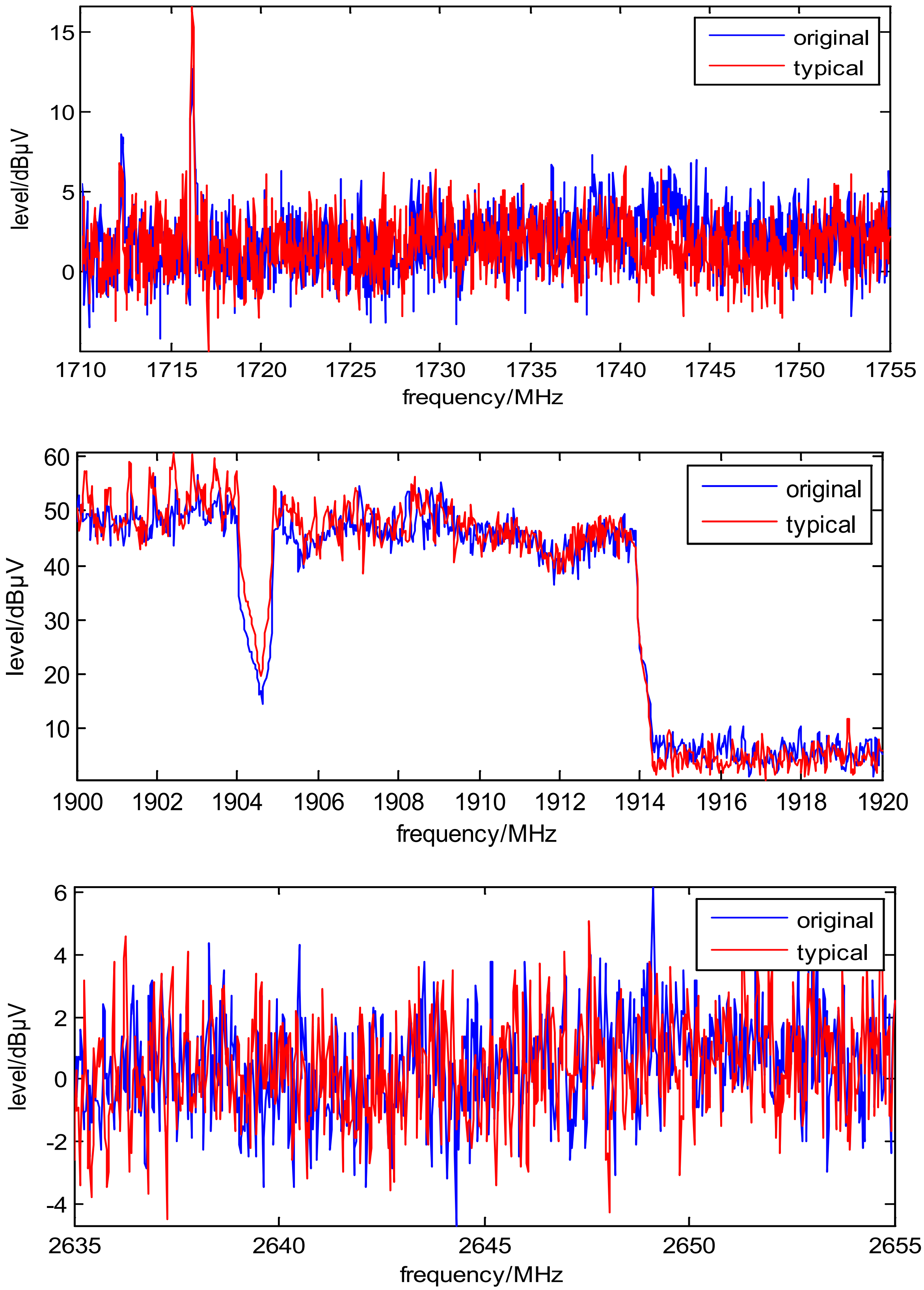


Fig.7. Comparison between original spectrum and typical spectrum.

According to these two types of files, we can evaluate the compression effect through three parameters: Compression ratio CR, Percent Root means square Difference PRD and standard Percent Root mean square Difference SRD, as shown in Table 1.

The formula of compression ratio (CR) is:

$$CR = \frac{D_{com}}{D_{orig}} \times 100\% \quad (4)$$

Where CR is compression ratio, D_{com} is the size of compressed spectrum data, D_{orig} is the size of original spectrum data.

The formula of the Percent Root means square Difference (PRD) is:

$$PDR = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - X_i)^2}{\sum_{i=1}^{n} Y_i}} \times 100\% \quad (5)$$

The formula of the standard Percent Root means square Difference (SRD) is:

$$SRD = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - X_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \times 100\% \quad (6)$$

Where Y_i and X_i is the i-th value corresponding to the original spectrum data and the compressed spectrum data, $\bar{Y}$ is the average of the level of the original spectrum.

It can be seen from table 1 and Fig7 that the overall compression efficiency of the algorithm is high, the CR of some frequency bands is close to 2%, and the PRD is close to 9%, especially in these frequency bands with less noise. At the same time, we also find that when the noise is dominant, R_{min} is generally small, and there are too many typical spectra that meet the conditions, So the error is relatively large.

In fact, in the similarity judgment stage, the most similar typical spectrum should be used to represent the original spectrum. Due to the limitation of computing resources, we only use serial number of the first qualified typical spectrum to replace the original spectrum in the sequential calculation. In this way, it may appear that the typical spectrum used to represent the original spectrum is not the best one.

6. Conclusion

The algorithm proposed in this paper can adaptively compress real-time spectrum data without serious distortion. It can be widely used in on-line and off-line compression of swept spectrum data in radio monitoring process, and does not need a priori knowledge. The algorithm replaces the spectrum with the serial number, so that the serial number can be transmitted in real time through the network while compressing, which greatly reduces the consumption of network bandwidth.

In future, we will further optimize the similarity calculation process to quickly match the best typical spectrum and reduce the compression error of the original spectrum. In addition, on this basis, we can further study the compression for storage and transmission based on coding to further improve the compression efficiency.

Acknowledgments

This work was supported in part by the key R&D Project implemented jointly by Sichuan and Chongqing in 2020 under Grant cstc2020jscx-cylhX0004.

References

[1] Alessio S M. Digital Signal Processing and Spectrum Analysis for Scientists[J]. Signals & Communication Technology, 2016.

[2] Kulin M, Kazaz T, Moerman I, et al. End-to-End Learning From Spectrum Data: A Deep Learning Approach for Wireless Signal Identification in Spectrum Monitoring Applications[J]. IEEE Access, 2018, 6:18484-18501.

[3] Man F, Shi R, He B. The data mining in wireless spectrum monitoring application[C]// IEEE International Conference on Big Data Analysis. IEEE, 2017.

[4] Attard R, Kalliovaara J, Taher T, et al. A high-performance tiered storage system for a global spectrum observatory network[C]// International Conference on Cognitive Radio Oriented Wireless Networks & Communications. IEEE, 2014.

[5] Skodras A N, Christopoulos C A, Ebrahimi T. JPEG2000: The upcoming still image compression standard[J]. Pattern Recognition Letters, 2001, 22(12):1337-1345.

[6] Chen S, A Be rmak, Wang Y, et al. Adaptive-Quantization Digital Image Sensor for Low-Power, Real-Time, Image Compression[J]. IEEE Transactions on Circuits & Systems I Regular Papers, 2007, 54(1):13-25.

[7] Tang Z, Peng T. A Wideband Spectrum Data Compression Algorithm base on Energy Detection[J]. Applied Mathematics & Information Sciences, 2015, 9(1):419-424.

[8] Li Y, Gao Z, Huang L, et al. A Wideband Spectrum Data Segment Compression Algorithm in Cognitive Radio Networks[C]// Wireless Communications & Networking Conference. IEEE, 2017.

[9] Wang K, Mi J, Xu C, et al. Real-time big data analytics for multimedia transmission and storage[C]// IEEE/CIC International Conference on Communications in China. IEEE, 2016.

[10] Elmore K L , Richman M B . Euclidean Distance as a Similarity Metric for Principal Component Analysis[J]. Monthly Weather Review, 2010, 129(3):540-549.

[11] Benesty J , Chen J , Huang Y . On the Importance of the Pearson Correlation Coefficient in Noise Reduction[J]. IEEE Transactions on Audio Speech and Language Processing, 2008, 16(4):757-765.

[12] Tegowski J, Deane G B, Lisimenka A, et al. Spectrum and statistical analyses of ambient noise[C]// Meetings on Acoustics. Acoustical Society of America, 2013:070079.

[13] Hammouda M A, Wallace J W. Noise uncertainty in cognitive radio sensing: Analytical modeling and detection performance[C]// International Itg Workshop on Smart Antennas. IEEE, 2012.

[14] Yu-Long Z, Geng-Xin Z, Jing H U, et al. Signal Detection Algorithm Based on Histogram[J]. Journal of Military Communications Technology, 2017.

[15] Kobayakawa M, Kinjo S, Hoshi M, et al. Fast Computation of Similarity Based on Jaccard Coefficient for Composition-Based Image Retrieval[M]// Advances in Multimedia Information Processing - PCM 2009. Springer Berlin Heidelberg, 2009:949-955.

[16] Zhang J N, Liu Z C, Pei Z. A New Method Based on Radio Frequency Spectrum for Background Noise Curve Extraction[J]. Applied Mechanics and Materials, 2013, 475-476:6.