Accurate identification on individual similar communication emitters by using HVG-NTE feature

Ke LI, Wei GE, Xiaoya YANG, and Zhengrong XU*

1School of Information and Computer, Anhui Agricultural University, Hefei, Anhui, 230036, China
2Anhui Provincial Engineering Laboratory for Beidou Precision Agriculture Information, Anhui Agricultural University, Hefei, Anhui, 230036, China
3Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Shanghai University, Shanghai, 200072, China

Abstract. Individual identification of similar communication emitters in complex electromagnetic environment has great research value and significance in both military and civilian fields. In this paper, a feature extraction method called HVG-NTE is proposed based on the idea of system nonlinearity. The shape of degree distribution, based on the extraction of HVG degree distribution, is quantified with NTE to improve the anti-noise performance. Then XGBoost is used to build a classifier for communication emitter identification. Our method achieves better recognition performance than the state-of-the-art technology on the transient signal data set of radio stations with the same plant, batch, and model, and is suitable for the case of small sample size.

Key words: communication emitter; identification; feature extraction; HVG; NTE.

1. Introduction

Individual identification of similar communication emitters refers to recognizing particular individuals of communication emitters of the same type[1]. In the military field, using prior information to identify which communication emitter of the enemy the signal comes from can provide an important basis for accurately predicting the enemy's strategic and tactical intentions [2, 3]. Meanwhile, in the civil field, this technology also has important application prospects in wireless communication network security [4], cognitive radio [5] and mechanical fault diagnosis [6].

When the transmitter is turned on or off or the communication mode is switched, the transmitted signal will undergo a transient process, called the transient signal [7]. The transient signal is the impulse response generated by each component, which contains rich fingerprint information [8,9]. Previous researches have mainly extracted transient signal features from instantaneous amplitude, phase and frequency [10, 11], such as wavelet analysis [12] and fractal analysis [13].

Dubendorfer et al. were the first to point out that individual identification of wireless transmitter emitters can be performed using time-frequency analysis methods [14]. Reising used discrete Gabor transform analysis to extract the statistical features of normalized amplitude coefficient sequences in the WiMax transient signal to identify individual emitters. In the case of SNR >-3db, the correct recognition rate exceeded 90%, but the experiment was carried out in an office environment, which could not meet the actual requirements [15]. Yuan used the Hilbert-Huang Transform [16] to obtain the time-frequency energy distribution of transient communication signals, but the process of obtaining the eigen mode function from HHT was inefficient, and there was a serious boundary effect [17].

Carroll was the first to point out that since the communication transmitter contains a semiconductor-based power amplifier, the transmitter system can be considered as a nonlinear system, and the RF fingerprint of the communication emitter can be extracted from the perspective of nonlinear dynamics [18]. Sun proposed a method based on singular spectrum analysis to analyze the transient signals of mobile phones, which is effective even in the case of low SNR [19]. But the experiment used different types of emitters, and did not use the same type of communication emitters. Jia takes different transmitters as chaotic systems and extracts one-dimensional component natural measures as subtle features, which can solve the problem of small data volume and low sampling rate [20]. But the SNR of all data samples is about 18db, which is not applicable under the condition of low SNR.

In this paper, we propose an individual identification method of communication emitters based on system nonlinearity. The method includes two stages: feature extraction and identification recognition. In the feature extraction stage, the improved phase method is used to detect the transient starting point, then the adaptive threshold method based on HHT energy trajectory is used to detect the transient end point, and finally the fingerprint features of communication emitters are obtained by the HVG-NTES feature extraction method to improve the anti-noise performance. In the identification and recognition

*e-mail: xzr_525@ahau.edu.cn
2. Proposed Method

Visibility graph (VG) [21] is a tool for analyzing non-linear time series and is a technique for mapping time series into graphs [22]. Horizontal visibility graph (HVG) [23] is a simplification of VG. Its method is simple and the calculation speed is faster than VG, so the horizontal visibility algorithm is more practical in the actual communication confrontation environment. The definition is as follows [24]:

Let \( \{ y_t \}_{t=1}^{T} \) be the time series of T data, and assign each data of the sequence to a node in the HVG. In the time series connecting \( y_i \) and \( y_j \), if a horizontal line can be drawn in the time series that does not intersect with any intermediate data height, the two nodes \( (t_i, y_i) \) and \( (t_j, y_j) \) in the graph are connected. Therefore, if the following geometric criteria are met within the time series, then \( (t_i, y_i) \) and \( (t_j, y_j) \) are two connected nodes:

\[
y_i, y_j > y_l
\]

Figure 1 shows an intuitive example of HVG, where the upper half draws a time series and the lower half represents the graph generated by the horizontal visibility algorithm. Each data in the sequence corresponds to a node in the graph. If the height of the data corresponding to the two nodes is greater than the height of all the data between them, the two nodes are connected.

![Fig.1. Illustration of the HVG](image)

We can extract the information of the graph structure after finding the graphs of the time series. Here we distinguish them by the degree distribution of the chaotic process [25]. For a given graph G with T nodes, the degree of node k is the number of edges associated with the node. The degree distribution \( P(k) \) is the proportion of nodes with degree k.

Because of the short duration, short data length and non-stationary characteristics of the transient signal of communication emitter, it is generally considered as a non-extensive system. In this paper, the non-extensive entropy of the degree distribution, Tsallis entropy (TE) [26], is used to quantify the shape of the degree distribution as a subtle feature of the communication emitter. For the discrete distribution \( P = \{ p_i, i = 1, 2, ..., T \} \), TE is defined as follows:

\[
S_q(P) = -\sum_{j=1}^{T} p_j^q \ln p_j, \quad (q \neq 1)
\]

Where \( \ln p_j = \frac{1-q}{1-q} \), the parameter q is a non-negative real number. The normalized TE is obtained as follows:

\[
H_q(P) = \frac{S_q(P) - S_{q,\text{min}}(P)}{S_{q,\text{max}}(P) - S_{q,\text{min}}(P)}
\]

The NTE calculated from the HVG degree distribution is expressed by Horizontal Visibility Graph-Normalized Tsallis Entropy (HVG-NTE). In this paper, the detected transient signal is regarded as a nonlinear time series. The first three are selected from the intrinsic mode functions (IMFs) obtained by empirical mode decomposition of the transient signals [27], which contain more nonlinear complexity components, and then the HVG-NTE is extracted. For each transient signal data, the following feature vectors can be obtained:

\[
v = [v_1, v_2, v_3]
\]

Where \( v_1, v_2, v_3 \) are the HVG-NTES obtained from the first three IMFs respectively.

3. Experimental Results and Discussion

3.1 Signal Acquisition. Figure 2 shows the schematic diagram of communication emitter signal receiving and collecting subsystem. In this experiment, 4 short-wave handheld radio stations with the model of QYT CB were used as the identification objects, and the communication frequency was 50MHz. The signal acquisition equipment is the Egrit RM200 receiver, the sampling frequency band is set at 49.75-50.25MHz, the sampling rate is 7KHz, and the radio frequency signal is converted to an intermediate frequency of 70KHz. The sampled data is stored in the host computer in hexadecimal data format. The host computer model is Lenovo ThinkCentre M910T, equipped with i7-7700 processor and 16G memory. It should be noted that
the experimental data was obtained under the condition of the out-field, considering the signal power and noise level interference, both the radio and the receiver use antennas to send and receive signals, rather than an ideal laboratory environment.

![Schematic diagram of communication emitter signal receiving and collecting subsystem](image)

**Fig. 2. Schematic diagram of communication emitter signal receiving and collecting subsystem**

### 3.2 Transient detection.

The typical transmission signal waveform is shown in Figure 3, including noise, transient signal and steady-state signal.

![Waveform of the transmission signal](image)

**Fig. 3. Waveform of the transmission signal**

The duration of the transient signal is very short. If there is an error in the detection of the starting and end points of the transient signal, it will affect the accuracy of identification recognition. We used an improved phase-based method to detect the transient starting point [28]. The difference between each successive phase variance is calculated twice to obtain the fractal trajectory (FT). At the beginning of the transient, the fractal trajectory should be zero, that is, the transient starting point is detected as the 90th frame, corresponding to the 1800th point in the original signal, as shown in Figure 4.

![Fractal trajectory](image)

**Fig. 4. Fractal trajectory**

![Energy trajectories based on HHT](image)

**Fig. 5. Energy trajectories based on HHT**

The adaptive threshold method based on energy trajectory detects the transient end point [29]. It can be seen from Figure 5 that after the end point $P_t$, positive slope value must exist reasonably as the energy is stable and almost equal. After the first point with a positive slope, the point where the energy trajectory is greater than the maximum value plus the standard deviation is the end point.

### 3.3 Robustness and anti-noise performance of NTE.

Here, the performance of the proposed NTE is mainly proved by Rossler map, and compared with Shannon entropy (SE). Rossler map [30] is given by

\[
\begin{align*}
\frac{dx}{dt} &= -z - y \\
\frac{dy}{dt} &= x + 0.15y \\
\frac{dz}{dt} &= 0.2 + K(zx - 5)
\end{align*}
\]

Four different dynamic systems were selected respectively, with parameters $K = 0.6, 0.7, 0.8$, and $0.9$, and the nonlinear complexity of the system increased continuously. The time series is obtained by integrating in increments of $0.01$, and the range of the parameter $q$ is set to $0.1$ to $1.5$, and $T$ is $1000$. The results are shown in Figure 6.
As can be seen from Figure 6, when there is no noise, SE can improve the recognition performance when increasing q value appropriately, but as q continues to increase, the discrimination of the dynamic system will decrease. NTE can clearly distinguish these four dynamic systems under different non-extensive coefficients q.

In order to evaluate the anti-noise performance of the two algorithms, the noise is superimposed on each time series by adding gaussian random variables with different noise levels.

**Figure 7** shows the entropy of the Rossler map of noise pollution (noise level is 0.2), and it can be seen that NTE has good resolution ability in the range of 0.1 ~ 1.5. However, no matter how much q is taken, SE can hardly distinguish these four different systems.

**Figure 8** shows the results of NTE and SE of four dynamic systems changes with noise level when q is 0.5. The graphs demonstrate the impact of different noise levels on the recognition performance of the two algorithms.
4. Fingerprint Feature Extraction

We extracted the HVG-NTE features from the transient signal of four handheld radio stations. In order to indicate the difference of fingerprint features between communication emitters, we visualized the feature distribution.

As can be seen from Figure 9, the feature sets of the four stations are significantly different, which indicates that the four stations can be distinguished by the method based on HVG-NTE.

5. Emitter Identification

XGBoost [31] generates a second-order Taylor expansion for the loss function, and obtains the optimal solution for the regular term outside the loss function, making full use of the parallel computing advantages of the multi-core CPU to improve the accuracy and speed. Therefore, we use XGBoost to establish an individual recognition classifier for communication emitters. We use the grid search method with the optimal parameters n_estimators=300, learning_rate=0.01, max_depth=20. 100 samples were collected from each handheld station, and the samples from each station were divided in a 1:1 scale, that is, 50 samples were taken as the training set, and the remaining 50 samples were taken as the test set. The experiment was performed 20 times independently and the average result was calculated.

6. Conclusion

This paper proposes a method for individual identification of communication emitters based on system nonlinearity. This method obtains IMFs of transient signals through empirical mode decomposition, extracts the degree
distribution of the first three IMFs from the received signals based on the HVG algorithm, and uses NTE to quantify the shape of the degree distribution as the fingerprint features of the communication emitters. The anti-noise performance and robustness of the fingerprint features are proved by Rossler map of different non-extensive coefficient \( q \) and noise level. Then a classifier based on XGBoost for individual identification of communication emitters is constructed. In this paper, the method of identifying radios with the same plant, batch, and model is implemented under the condition of small sample size and low SNR, whose performance is better than the state-of-the-art method.

**Acknowledgments.** This work was supported by the Provincial-level Natural Science Research Key Project of Colleges and Universities in Anhui Province. (No. KJ2018A0145)

**REFERENCES**


