International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.2 (2016), pp.127-134
http://dx.doi.org/10.14257/ijsip.2016.9.2.11

Identifying of Digital Signals Based on Manifold Learning

Qingbo Ji[1], Boyang Feng[1], Yun Lin[1*], Zheng Dou[1*], Zhiqiang Wu[2] and Zhiping Zhang[2]

[1]College of Information and Communication Engineering
Harbin Engineering University, Harbin, China
linyun@hrbeu.edu.cn
[2]Department of Electrical Engineering
Wright State University, Dayton, Ohio, U.S.
zhiqiang.wu@wright.edu

Abstract

Modulation type is one of the most important characteristics used in signal recognition. An algorithm to realize signal modulation identification is proposed in this paper. We applied wavelet transformation and STFT to the signal, and then used manifold learning method to reduce the high dimension and extracted the recognition feature. The proper threshold value was set as the classifier to achieve the purpose of recognizing 4 kinds of signals (MASK, MFSK, MPSK,QAM) in Gauss white noise environment. The algorithm requires priori signal information no other than signal-to-noise rate. Simulation result indicates the algorithm achieves good performance.

Keywords: *Digital signals identification, Feature extraction, Manifold learning method, Isomap*

1. Introduction

Digital signals are widely used both in commercial and military fields. To analysis the information transferred by the source, the signal mode needs to be figured out by selecting the appropriate features. Feature selection is the process of choosing a subset of the original predictive variables by eliminating redundant and uninformative ones. By extracting as much information as possible from a given data set while using the smallest number of features, we can save significant computing time and often build models that generalize better to unseen points.

Among all the signal parameters, in-pulse characteristics have very special effects. Many in-pulse characteristics have been used on signal recognition such as entropy analysis, short time Fourier transformation, wavelet transformation, complexity feature and so on. For example, Swami and Sadler [3] proposed a wavelet transform-based signal identification method, with which the success rate of 98% at signal-to-noise ratio (SNR) 4 dB was reported. Zhang [4] proposed a support vector machine-based classifier to classify the signals according to the proposed features. The types of the signals have been identified with a success rate of about 90% for 0<SNR<5 dB. A digital modulation classification system was proposed by Xu et al. [5] for CR using only temporal waveform features. They reported a success rate of 95% at SNRs ranging from 10 to 80 dB. In [6], the authors presented a high-performance multi-layer perception neural network with resilient back propagation learning algorithm. In [7], a signal classification approach based on neural network ensembles was proposed, which enables dynamic spectrum access. From the research works mentioned above, it can be found that: (a) most of the proposed methods can only recognise low-order and limited digital signals; (b) most of

* Corresponding Author

ISSN: 2005-4254 IJSIP

methods require high SNRs; (c) machine learning-based methods may have higher performance.

This paper focuses on the study of feature extraction part to realize the identification in lower SNRs. Manifold learning[1,2] procedures can realize the visualization by embedding the high-dimension data into 2 or 3dimensions while preserving as much as possible the metric in the natural feature space, which makes observation and analysis easier. Recently, manifold learning algorithms are increasingly used in intelligent cognitive system [8-11]. New types of signals can be automatically detected, classified, and identified in a cognitive environment. The manifold learning algorithms overcome the limitations of existing linear methods such as principal component analysis (PCA) and independent component analysis [12]. They have been successfully applied in signal and image processing and pattern recognition.

2. Isomap Method

To commence, suppose R is a nonempty set [13]. The whole topological space is defined by a set of topologically equivalent objects. A manifold M is a topological space that is locally Euclidean, i.e. There is a neighborhood around every point of M that is topologically the same as the open unit ball in R^d , so M is a d- dimensionality topological manifold. In general, any object that is nearly flat on small scales is a manifold. An open line segment, a circle and a knotted circle are 1-manifold (d = 1) that are mapped in one-, two- or three-dimensional space, respectively. This means that, although the mapping spaces of these samples are different, but they have similar intrinsic dimensions.

The Isomap algorithm is based on multidimensional scaling (MDS). The data is mapped from a high-dimensional input space to the low-dimensional space of a nonlinear manifold depending on global invariants. Follow the steps outlined in Figure 3.

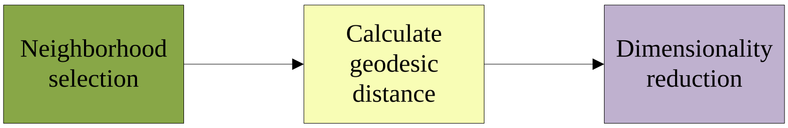

Figure 3. Isomap Method

(1) Find the k-nearest neighbor or ε - neighborhood of each point of data space $\{x_i\}_{i=1}^{l}$. Where k represents the number of points chosen, or ε is the area radius. Concatenate each point with its neighborhood to constitute the proximity graph. Use the Euclidean distance as the weight of each edge.

(2) Calculate the geodesic distance of the data in the proximity graph as the shortest distance using standard graph search methods like Dijkstra's algorithm and Floyd's algorithm. The latter one is chosen because it fits more to the computer simulation. The basic idea of Floyd's algorithm is that, there are no more than two ways to find the shortest path between point A and B, from A directly to B or through several points. Assume *dis(AB)* is the shortest distance between point A and B, X is arbitrary point. If $dis(AX)+dis(XB)<dis(AB)$, $A \to X \to B$ is the shortest path.

(3) Reduce the dimension with the classical metric MDS using the geodesic distance. Assume Y is a d-dimensionality space, $y_i (i=1,2,\cdots N)$ is the coordinate vector of the points in Y, *error* is the loss during the embedding.

$$error = \left\| \tau(D_G) - \tau(D_Y) \right\|_{L^2} \tag{1}$$

Where $D_G = \{d_G(i,j) = \|y_i - y_j\|_{L^2}\}$ represents the Euclidean distance matrix of high- dimensionality, while $D_Y = \{d_Y(i,j) = \|y_i - y_j\|_{L^2}\}$ is the same matrix of low-dimensionality, and the operator τ transforms the calculation of distance into inner product operation.

$$\tau(D) = -\boldsymbol{HSH} / 2 \tag{2}$$

Where $S = \{S_{ij} = D_{ij}^2\}$ is square distance matrices, $H = \{H_{ij} = \delta_{ij} - 1/N\}$ is center matrices.

Our procedure should minimize error. Presume λ_i is the ith eigenvalue of $\tau(D_G)$ with y_i being the corresponding ith eigenvector. Sort λ_i from the lowest to the highest, y_i rearranged with it. The element in row i and column j of Y is $\sqrt{\lambda_i}\, y_{ij}$.

To obtain the feature of set $X = \{x_k\}_{i=1}^{l} \subset R^n$ in n-dimensionality space, we need to estimate the intrinsic dimensionality $\hat{d}$ and the optimal neighborhood size $\hat{k}$. If d is valued too small, the disconnected parts would be mapped into one area; if it is overvalued, the manifold would contain too much redundant information. Once k is valued too small, the entire data set would be mapped into a local neighborhood rather than global mapped; while if it is too small, imagine the points ought to be mapped into one area separated apart, the manifold would be obviously false without representing the global property of the original data.

The intrinsic dimensionality $\hat{d}$ can be obtained by the drawing curve of the error shown in Figure 4. Tenenbaum purposed a method estimating the optional $\hat{d}$ of Isomap[2] with finding the “elbow” of the error curve, where the curve stops falling sharply. According to Fig. 4, the error is small enough to maintain the data integrity when $d \le 2$. We choose $\hat{d} = 2$ in order to visualize the result, make the progress easier to analyze.

Experiment results indicate that the manifold of the data is always integrated when k=8, so we choose the optimal neighborhood size $\hat{k} = 8$.

3. Recognition

3.1. Signals Representation

A digitally modulated signal can be represented as

$$s(t) = \mathrm{Re}\left\{A(t)g(t)e^{(j2\pi f(t)+\phi(t))}\right\} \tag{1}$$

In which $A(t)$ is the amplitude, $g(t)$ is the response of the symbol pulse shaping filter, $f(t)$ is the carrier frequency and $\phi(t)$ is the phase. FSK, ASK, PSK and QAM signals can be represented as followed:

$$\mathrm{S_{MFSK}}(\mathrm{t}) = \frac{\sqrt{2E_S}}{T_S}\mathrm{Re}\left\{\sum_k e^{j2\pi(f_c+\Delta f_k)t} g(t-kT_S)\right\} \tag{2}$$

in which $\Delta f_k = \left[k - \left(M_s - 1/2\right)\right]\Delta f, k = 0,1,\cdots,M_s - 1$, $g(t)$ is the pulse shaping function, f_c is the carrier frequency, M_s is the number of states, T_s is symbol period and period and E_s is energy per symbol, Δf_k denotes the symbols.

$$\mathrm{S_{MASK}}(\mathrm{t}) = \frac{\sqrt{2E_S}}{T_S}\mathrm{Re}\left\{\sum_k A_k e^{j2\pi f_c t} g(t-kT_S)\right\} \tag{3}$$

In which $A_k = 2k - M_s - 1, k = 0,1,\cdots,M_s - 1$, A_k denotes the symbols.

$$\mathrm{S_{MPSK}}(\mathrm{t}) = \frac{\sqrt{2E_S}}{T_S}\mathrm{Re}\left\{\sum_k \Gamma_k e^{j2\pi f_c t} g(t - kT_S)\right\} \tag{4}$$

In which $\Gamma_k = e^{j2\pi f_c t}, k = 0,1,\cdots,M_s - 1$.

$$S_{\mathrm{MQAM}}(t) = \sum_{i=1}^{N}(A_i + B_i)u_T(t - iT),$$
$$A_i, B_i \in \{2k - 1 - K, k = 1,2,\cdots,K\} \tag{5}$$

Where $u_T(\cdot)$ is unit impulse with period T.

3.2. STFT Manifold

The short time Fourier transform (STFT) expression of the signal can be expressed as (6)

$$STFT_x(t,f) = \int_{-\infty}^{\infty} x(\tau)g^*(\tau - t)e^{-j2\pi f\tau}d\tau$$
(6)

Where $x(\tau)$ is the signal, $g(t)$ is the window function, $g^*(t)$ is the conjugate of $g(t)$. The main idea of the STFT is to add window function to signal. Assume that the signal within the window length is stationary, then fourier transforming the windowed signal. The length of the window function directly affects the signal resolution in time domain and frenquency domain, consequently, the recognition result of radio signals are affected. Select proper length of the window function can achieve good identification effect.

As shown in Figure 1, the different signals Isomap embedding varies in distribution obviously, so we can identify the signal type by manifold variance feature (MVF). In this case, threshold value method is available as the classifier, which will be represented in chapter 4.

3.3. WT Manifold

WT is chosen to analysis the signal because the signal WT domain contains the information both in time and frequency domain, and it seems to be less influenced by noise. The wavelet transform of signal $f(t)$ is defined as following.

$$W_f(a,b) = \langle f, \psi_{a,b} \rangle = \int_{-\infty}^{+\infty} f(t)\psi^*_{a,b}(t)\mathrm{d}t \quad , a \neq 0 \tag{5}$$

Where a represents the scale parameter, b represents the translation parameter (time shifting), and the basis function $\psi_{a,b}(t)$ is obtained by scaling the mother wavelet $\psi(t)$ at time b and scale a,

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a})$$

For each scale, we get a set of data, so the result is a-dimension matrix. We need to extract the basis information from the intricate data, as introduced in next part.

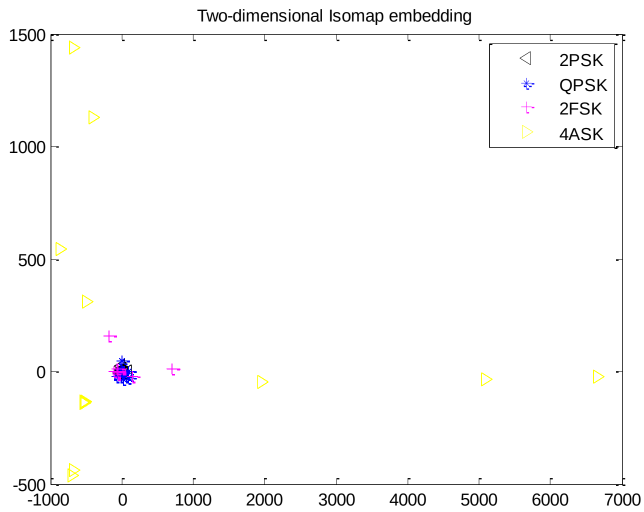

Figure 1. 4 Digital Signals Isomap Embedding

3.4. Signal Identification

We realized the signal identification as the processer followed shown in Figure 2:

(1) After STFT, high-dimension data is obtained. The signal can be roughly identified by the MVF. If the signal belongs to MQAM, then we have to analysis it with the second step. If not, the signal type is already identified.

(2) To figure out the received signal is 16QAM, 64QAM or 128QAM, WT is applied to it. The MVF indicates the modulation of the signal obviously this time.

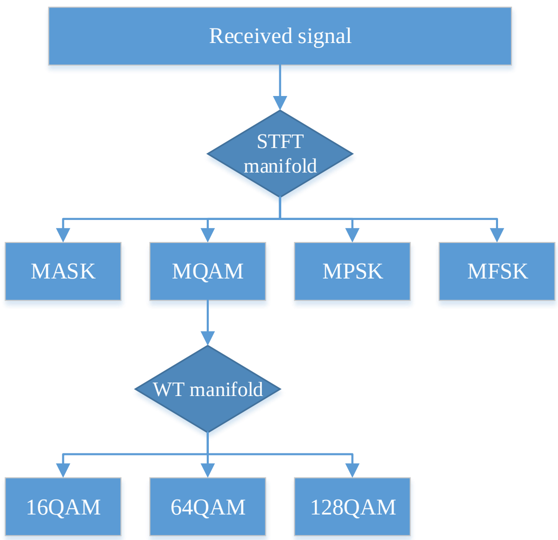

Figure 2. Basic Flow of the Signal Identification System

4. Simulation

In this section, we provide some experiments for our methods. Using the algorithm proposed, the identification of 4ASK, 2PSK, QPSK, 2FSK, 16QAM, 64QAM, 128QAM are carried out in the MATLAB simulation platform, and simulation results show the effectiveness of the algorithm.

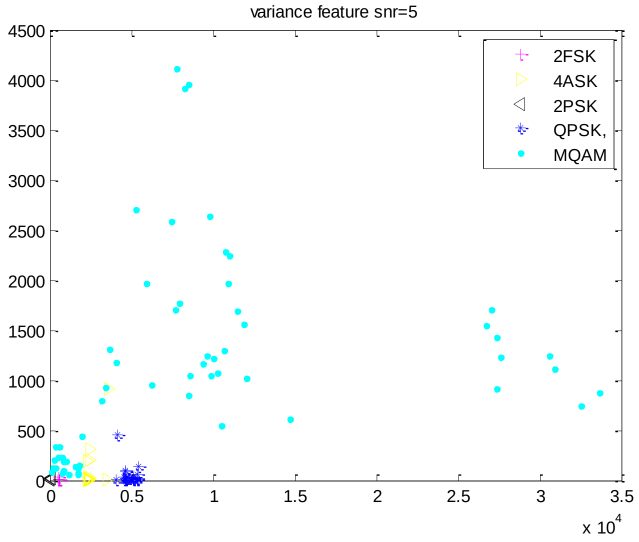

Figure 3. STFT Manifold Variance Feature of Digital Signals

Simulation condition is: signal carrier frequency is 0.1MHz; sampling frequency is 20MHz; the range of signal-to-noise rate is from -10~5 dB.

An experiment result is shown in Figure 3. The STFT MVF of 5 kinds of signals can be separated clearly, so we choose appropriate thresholds as the classifier of signals type. But MQAM (16 QAM, 64 QAM, 128QAM) signals can not be identified by this method. To solve this problem, we apply WT manifold method to MQAM signals. The result is shown in figure 4.

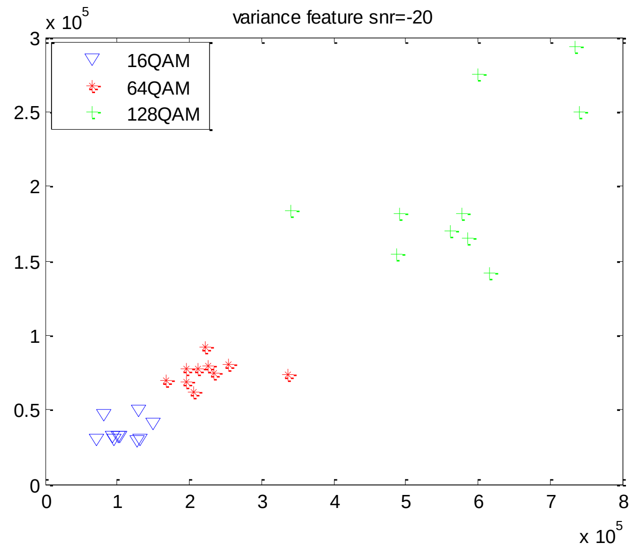

Figure 4. WT Manifold Variance Feature of MQAM Signals

100 experiments are applied to get the range of each kind of signals referred. With the noise influencing, the thresholds would be vague, so the recognition rate would fall sharply at a specific point.

To compare our method with the SIEMAP method [14], the recognition rate is shown in Figure 5 and 6. Our method is obviously more effective in lower SNR environment, when SNR>-5dB, the recognition rate is nearly 100%, which implied the good performance of the method based on STFT and WT manifold.

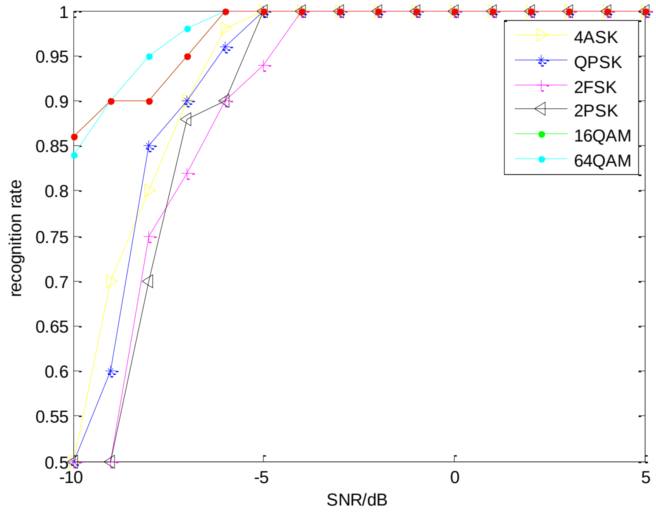

Figure 5. Recognition Rate of STFT Combing with WT Manifold

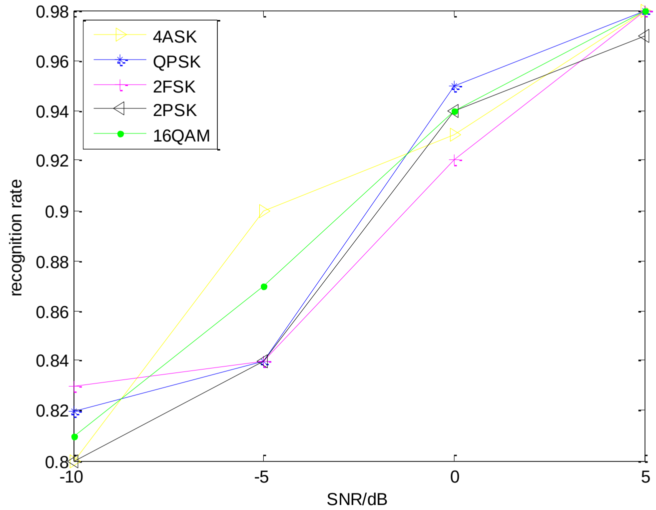

Figure 6. Recognition Rate of Method based on SIEMAP

4. Conclusion

In this paper, we proposed a digital signals identification method based on manifold learning method. We combine STFT and WT with Isomap method, then extract the variance feature as the identify feature to achieve our goal. We applied our method to MASK, MPSK, MFSK and MQAM signals. The simulation shows when SNR>-5dB, the recognition rate is nearly 100% which proves the validity of the method in very low SNR condition.

Compared with other methods, this method does not require high sampling rate, it achieves better recognition rate in Gauss white noise environment. The disadvantage of the method is that, it takes more time to reach the high percentage of correct identification, this problem would surely be solved with the development of the hardware and related software. In addition, a more efficient classifier might be used for improving the recognition rate.

Acknowledgments

This work was supported by the Key Development Program of Basic Research of China(JCKY2013604B001), the Nation Nature Science Foundation of China (No.61301095 and No.61201237), Nature Science Foundation of Heilongjiang Province of China (No. QC2012C069, F201408 and F201407) and the Fundamental Research Funds for the Central Universities (No. HEUCF1508).

Meantime, all the authors declare that there is no conflict of interests regarding the publication of this article.

References

[1] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding", Science, 290, 5500 **(2000)**.

[2] J. B. Tenenbaum V. D. Silva, J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction", Science, vol. 290, no. 5500, **(2000)**.

[3] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants", IEEE Trans. Commun., vol. 48, no. 3, **(2000)**.

[4] Q. Zhang, "Cognitive radio on a reconfigurable MPSoC platform", Report of University of Twente **(2009)**.

[5] Y. Xu, L. Ge and B. Wang, "Digital modulation recognition method based on tree-structured neural networks", Int. Conf. on Communication Software and Networks, Macau, China. **(2009)** February.

[6] A. Attar, O. Holland, M. R. Nakhai and A. H. Aghvami, "Interference-limited resource allocation for cognitive radio in orthogonal frequency division multiplexing networks", IET Commun. vol. 2, no. 6, **(2008)**.

[7] L. Bixio, M. Ottonello, H. Sallam, M. Raffetto and Regazzoni, "Signal classification based on spectral redundancy and neural network ensembles", Fourth Int. Conf. on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM) Hannover, Germany, **(2009)** June.

[8] A. N. Mody, S. R. Blatt, D. G. Mills, T. P. McElwain and J. D. Thammakhoune, "Recent advances in cognitive communications", IEEE Commun. Mag, vol. 45, no. 10, **(2007)**.

[9] X. Chen and J. Zhang, "Maximum Variance Difference Based Embedding Approach for Facial Feature Extrication", International Journal of Pattern Recognition and Artificial Intelligence, vol. 7, no. 24, **(2010)**.

[10] G. Peyre, "Manifolds for signals and images", Computer Vision and Image Understanding, vol. 209, no. 113, **(2009)**.

[11] A. Ebrahimzadeh and Ranjbar, "Intelligent digital signal-type identification", Artif. Intell., vol. 21, no. 4, **(2008)**.

[12] J. Wang, Z. Zhou and A. Zhou, "Machine learning and it's application", Tsinghua University Press, **(2006)**.

[13] J. R. Munkres, "Analysis on Manifolds", Science press, Beijing **(2012)**.

[14] S. Li, X. Wang and J. Wang, "Manifold learning-based automatic signal identification in cognitive radio neetworks", IET Communicitions, vol. 6, no. 8, **(2010)**.

[15] Z. Ji, W. Wang, C. Liu, "Identification of Radio Signal Pulse Modulation Characteristics", ICSP Proceedings, **(2012)**.

[16] C. Bao-yuan, L. Wei-ming, "The Algorithm Research of Speech End point Detection and Simulated Based on the Wavelet Analysis", Journal of Harbin University of Science and Technology, vol. 14, no. 1, **(2009)**.

[17] Z. Hui, Z. Dan-feng, H. Chang-bo and Z. Tie-lin, "Design and Implementation of MSK Modem", Journal of Harbin University of Science and Technology, vol. 16, no. 1, **(2011)**.

