Feature Extraction Methods for Signal Analysis: A Comprehensive Review

Abhishek Jain[1], Suhel Ahamed [2]
[1,2]Department of IT, Central University (GGU) Bilaspur Chhattisgarh India

Abstract:

Signal processing plays a crucial role in various domains, including telecommunications, biomedical engineering, audio processing, and image analysis. Feature extraction is an essential step in signal analysis, aimed at capturing relevant information from raw signals for further analysis and decision-making. This research paper provides a comprehensive review of recent feature extraction methods for signal analysis. It explores different techniques and algorithms used for extracting discriminative and informative features from signals. The paper also discusses the applications, strengths, and limitations of these methods, along with potential future research directions in the field of feature extraction for signal analysis.

1 Introduction

1.1 Background

Signal processing involves the analysis and manipulation of signals to extract meaningful information. Feature extraction is a crucial step in signal analysis, as it helps in reducing the dimensionality of signals and capturing essential characteristics for subsequent analysis tasks[1,6,7,15].

1.2 Problem Statement

The paper aims to review and analyze recent advancements in feature extraction methods for signal analysis. It seeks to identify the strengths, limitations, and potential applications of these methods, providing a comprehensive understanding of their utility in different domains.

1.3 Objectives

The objectives of this research paper are:

- To review and analyze recent feature extraction methods for signal analysis.
- To explore the underlying principles and algorithms used in these methods.
- To discuss the applications and potential benefits of feature extraction in different domains.
- To evaluate the strengths and limitations of various feature extraction techniques.
- To identify potential future research directions and challenges in the field.

2 Feature Extraction Techniques

2.1 Time-Domain Features

Time-domain characteristics, including as mean, variance, skewness, and kurtosis, extract information directly from the time-series signal. Other time-domain metrics that provide light on signal properties are the zero-crossing rate, energy, and entropy [2,10,13,14].

1. Mean: The average value of the signal over a given time interval.
2. Standard Deviation: A measure of the dispersion or spread of the signal values around the mean.
3. Variance: The average of the squared differences from the mean.
4. Skewness: A measure of the asymmetry of the signal distribution.
5. Kurtosis: A measurement of the signal distribution's peak or flatness.
6. RMS stands for root mean square, which is the average of squared signal values.
7. Crest Factor: The proportion of the signal's peak value to its RMS value.
8. The difference between the greatest and smallest signal values is known as the peak-to-peak amplitude.
9. The speed at which the signal crosses the zero axis or changes signs.
10. Time entropy: A gauge of the signal values' randomness or unpredictability.

11. A measurement of how similar a signal is to a time-delayed version of itself is called autocorrelation.
12. Energy is the sum of a signal's power or energy over a specific time period.
13. Maximum Value: The maximum signal value during the course of a certain time period.
14. Lowest Value

2.2 Frequency-Domain Features

Using methods like the Fourier transform or the Wavelet transform, frequency-domain analysis entails converting signals into the frequency domain. Spectral centroid, spectral flatness, and power spectral density are examples of frequency-domain characteristics that offer details on the distribution of signal energy across various frequencies[3,11,14,16].

1. Power Spectral Density (PSD): The distribution of power over different frequency components of a signal.
2. The weighted average of the frequencies in the signal, or the "centre of mass" of the spectrum, is known as the "spectral centroid."
3. A measurement of the complexity or unpredictability of the spectral content is called the spectral entropy.
4. A measurement of the spectrum's flatness or narrowband width.
5. The frequency below which a specific portion of the overall signal strength is present is known as the spectral roll-off.
6. Spectral flux is a measure of how quickly the spectrum changes over time.
7. A measure of the asymmetry in the spectral distribution is called spectral skewness.
8. A measurement of the peakedness or flatness of the spectral distribution is called spectral kurtosis.
9. Harmonic Ratio: The ratio of the energy in the harmonics to the total signal energy.
10. Fundamental Frequency: The lowest frequency component or the dominant frequency in the signal.
11. Total Harmonic Distortion (THD): The ratio of the root mean square (RMS) of the harmonics to the RMS of the fundamental frequency.
12. Band Energy Ratios: The ratio of energy in different frequency bands, such as low-frequency, mid-frequency, and high-frequency bands.
13. Frequency Variance: A measure of the spread or variability of the frequency content.
14. Peak Frequency: The frequency with the highest magnitude or power in the spectrum.
15. Bandwidth: The range of frequencies over which a specified percentage of the total signal power lies.

2.3 Statistical Features

Statistical features quantify the statistical properties of signals, such as correlation, autocorrelation, and covariance. These features help capture dependencies and relationships within the signal, enabling effective analysis and classification[4,5,18].

1. Mean: The average value of the signal.
2. Median: The middle value of the sorted signal.
3. Mode: The most frequently occurring value in the signal.
4. Standard Deviation: A measure of the dispersion or spread of the signal values around the mean.
5. Variance: The average of the squared differences from the mean.
6. A measure of the asymmetry in the signal distribution is called skewness.
7. Kurtosis: A measurement of the signal distribution's peak or flatness.
8. Range: The discrepancy between the signal's highest and lowest values.
9. The interval between the signal's 25th and 75th percentiles is known as the interquartile range (IQR).
10. Percentiles are the points below which a specific proportion of the signal values fall.
11. Coefficient of Variation: The ratio of standard deviation to mean, which shows how variable the signal is in comparison to other signals.
12. A measurement of how closely two signals resemble one another when they are delayed in time.
13. Cross-correlation: A measurement of the resemblance between two distinct signals occurring at dissimilar time delays.
14. Entropy: A gauge of the randomness or uncertainty in the signal values.
15. The whole amount of energy, or
16. Peak-to-Average Ratio (PAR): The ratio of the peak value to the average value of the signal.
17. Hurst Exponent: A measure of the long-term memory or self-similarity of the signal.
18. Fractal Dimension: A measure of the complexity or self-similarity of the signal.

2.4 Waveform Shape Features

Waveform shape features describe the shape or morphology of signals. These features can be extracted using techniques like morphological operations, contour analysis, or higher-order statistical analysis[8,12,17]. Examples of waveform shape features include peak amplitude, slope, duration, and curvature.

1. Rise Time: The time taken for the signal to rise from a specified lower threshold to a specified upper threshold.
2. Fall Time: The time taken for the signal to fall from a specified upper threshold to a specified lower threshold.
3. Peak Amplitude: The maximum amplitude or value reached by the signal.
4. The difference between the signal's maximum and minimum amplitudes is known as peak-to-peak amplitude.
5. Overshoot: The degree to which a signal deviates from its steady-state value or the desired value's target value.
6. Undershoot: The amount by which the signal goes below its steady-state value or falls short of the final desired value.
7. Pulse Width: The duration of a pulse or the width of a specific portion of the signal.
8. Duty Cycle: The ratio of the pulse width to the total period or duration of a periodic signal.
9. Period: The time taken for one complete cycle of the signal.
10. Frequency: The number of cycles or oscillations per unit of time.
11. Amplitude Modulation (AM) Depth: The extent to which the amplitude of a carrier signal is modulated by a modulating signal.
12. Phase Modulation (PM) Deviation: The extent to which the phase of a carrier signal is modulated by a modulating signal.
13. Harmonic Distortion: The presence of additional frequency components in the signal that are not present in the original input.
14. Crest Factor: The proportion of a signal's peak to RMS amplitude.
15. A measure of the asymmetry in the signal distribution is called skewness.
16. Kurtosis: A measurement of the signal distribution's peak or flatness.
17. Envelope: A curve that represents the variation of the signal's amplitude over time.
18. Shape Factor: A measure of the shape or profile of the signal waveform.
19. Symmetry Factor: A measure of the symmetry or balance of the signal waveform.

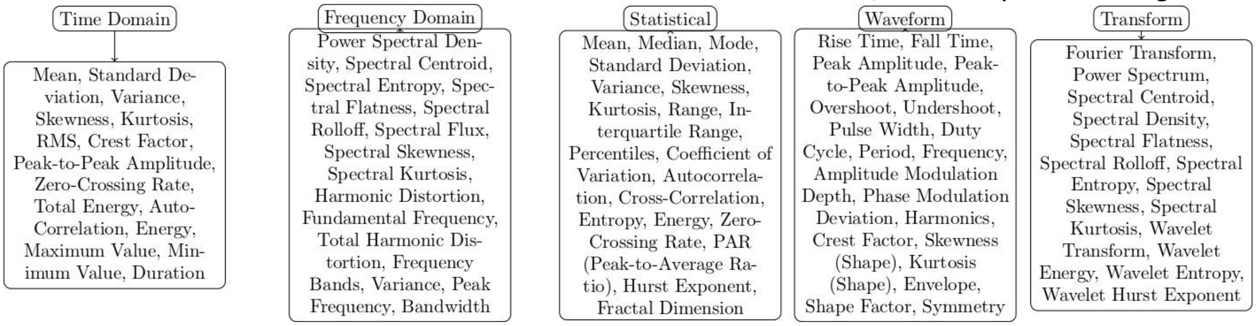
2.5 Transform-Based Features

Transform-based feature extraction methods utilize various signal transforms, such as Wavelet Transform, Discrete Cosine Transform, or Singular Value Decomposition, to extract informative features. These methods provide a compact representation of the signal, capturing relevant information in a transformed domain [1,5,9,11,19].

1. Fourier Transform: A mathematical transform that decomposes a signal into its constituent frequencies.
    - Power Spectrum: The distribution of power (magnitude squared) of the signal across different frequencies.
    - Spectral Centroid: The center of gravity of the power spectrum, indicating the average frequency content of the signal.
    - Spectral Density: The power spectrum normalized by the frequency resolution or bandwidth.
    - Spectral Flatness: A measure of the spectral richness or tonality of the signal.
    - Spectral Roll-off: The frequency below which a specified percentage of the signal's energy is contained.
    - Spectral Entropy: A measure of the spectral randomness or complexity of the signal.
2. Wavelet Transform: A mathematical transform that decomposes a signal into its constituent wavelet functions.
    - Wavelet Energy: The total energy or power of the signal across different wavelet scales.
    - Wavelet Entropy: A measure of the wavelet coefficient distribution's randomness or complexity.
    - Wavelet Packet Energy: The total energy or power of the signal across different wavelet packet nodes.
3. Short-Time Fourier Transform (STFT): A variation of the Fourier Transform that provides time-varying frequency analysis of a signal.
    - Spectrogram: A visual representation of the magnitude or power of the signal over time and frequency.
    - Mel Frequency Cepstral Coefficients (MFCCs): A feature representation commonly used in speech and audio processing.
4. Discrete Wavelet Transform (DWT): A variation of the Wavelet Transform that provides multi-resolution analysis of a signal.
    - Wavelet Coefficients: The coefficients obtained from the DWT decomposition at different scales and positions.
    - Wavelet Packet Coefficients: The coefficients obtained from the DWT decomposition using a wavelet packet basis.
5. Empirical Mode Decomposition (EMD): A data-driven decomposition method that decomposes a signal into intrinsic mode functions (IMFs).
    - IMF Energy: The energy or power of each IMF component.
    - IMF Frequency: The frequency content or dominant oscillations of each IMF component.

- IMF Instantaneous Frequency: The time-varying frequency of each IMF component.
- IMF Hjorth Parameters: Features derived from the IMFs, including activity, mobility, and complexity.

6. Discrete Cosine Transform (DCT): A transform similar to the Fourier Transform that represents a signal in terms of its frequency components.
   - DCT Coefficients: The coefficients obtained from the DCT, commonly used in image and video compression.



3 Applications and Case Studies

3.1 Biomedical Signal Analysis [1]

Feature extraction plays a vital role in biomedical signal analysis, such as electrocardiogram (ECG) analysis, electromyography (EMG) analysis, and electroencephalogram (EEG) analysis. The paper explores the use of different feature extraction methods in these domains and their impact on diagnosis and monitoring.

3.2 Audio Processing and Speech Recognition [6]

Feature extraction methods are widely employed in audio processing and speech recognition systems. Mel-frequency cepstral coefficients (MFCCs) and linear predictive coding (LPC) are commonly used techniques for extracting features from audio signals. The paper discusses the application of these methods and their effectiveness in speech recognition tasks.

4 Challenges and Future Directions

4.1 Handling Noisy and Variable Signals [9]

Real-world signals often contain noise, variability, and artifacts. Future research should focus on developing robust feature extraction methods that can effectively handle noisy and variable signals.

4.2 Multimodal Signal Analysis [13,20]

With the increasing availability of multimodal data, integrating feature extraction methods across different signal modalities presents an interesting research direction. Developing fusion techniques that combine features from different modalities can enhance the overall analysis and decision-making process.

4.3 Deep Learning Approaches [14,5,20,21]

Convolutional neural networks and recurrent neural networks, two examples of deep learning approaches, have demonstrated outstanding performance in a variety of signal processing applications. Exploring deep learning-based feature extraction methods and their integration with traditional techniques can further enhance the effectiveness of signal analysis.

4.4 Explainability and Interpretability [12,9,3,2]

As feature extraction methods become more complex, ensuring the interpretability and explainability of extracted features becomes crucial. Future research should focus on developing techniques that provide meaningful explanations for the extracted features, enabling users to understand and trust the analysis results.

5 Conclusion

This study included a thorough overview of signal analysis feature extraction techniques. . It discussed various techniques, algorithms, and applications of feature extraction in different domains. The strengths, limitations, and challenges of these methods were analyzed, along with potential future research directions. Feature extraction plays a vital role in signal analysis and enables effective analysis, classification, and decision-making in diverse fields.

References:

[1] R. Sinha, 'An Approach for Classifying ECG Arrhythmia Based on Features Extracted from EMD and Wavelet Packet Domains', 2012.

[2] R. Zhao, B. Du, L. Zhang, and L. Zhang, 'Beyond Background Feature Extraction: An Anomaly Detection Algorithm Inspired by Slowly Varying Signal Analysis', IEEE Transactions on Geoscience and Remote Sensing, 2016.

[3] S. Siuly and Y. Li, 'Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification', Computer Methods and Programs in Biomedicine, 2015.

[4] G. Rutkowski, K. Patan, and P. Leśniak, 'Comparison of Time-Frequency Feature Extraction Methods for EEG Signals Classification', International Conference on Artificial Intelligence and Soft Computing, 2013.

[5] A. Mane, P. S. D. Biradar, and P. R. K. Shastri, 'Review paper on Feature Extraction Methods for EEG Signal Analysis', 2015.

[6] M. Suchetha, N. Kumaravel, and B. Benisha, 'Denoising and arrhythmia classification using EMD based features and neural network', International Conference on Cryptography, Security and Privacy, 2013.

[7] C. Uyulan and T. T. Erguzel, 'Analysis of Time -- Frequency EEG Feature Extraction Methods for Mental Task Classification', International Journal of Computational Intelligence Systems, 2017.

[8] F.-Y. Leu, W. Caesarendra, and T. Tjahjowidodo, 'A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing', 2017.

[9] T. Uktveris and V. Jusas, 'Comparison of Feature Extraction Methods for EEG BCI Classification', International Conference on Information and Software Technologies, 2015.

[10] A. Hazarika, L. Dutta, M. Boro, M. Barthakur, and M. Bhuyan, 'An automatic feature extraction and fusion model: application to electromyogram (EMG) signal classification', International Journal of Multimedia Information Retrieval, 2018.

[11] J. Zabalza et al., 'Novel Two-Dimensional Singular Spectrum Analysis for Effective Feature Extraction and Data Classification in Hyperspectral Imaging', IEEE Transactions on Geoscience and Remote Sensing, 2015.

[12] F. Ulloa, 'Bearing fault detection through machine learning: time-domain vs time-frequency analysis for feature extraction', First EAGE Workshop on High Performance Computing for Upstream in Latin America, 2018.

[13] A. Matsuyama, M. Jonkman, and F. D. Boer, 'Improved ECG signal analysis using wavelet and feature extraction', Methods of Information in Medicine, 2005.

[14] B. Dumitrascu, N. Nistor, and D. Aiordachioaie, 'Analysis of transient signals by feature extraction from time-frequency images', International Symposium for Design and Technology in Electronic Packaging, 2017.

[15] A. Doulah, S. A. Fattah, W.-P. Zhu, A. Mo, and M. O. Ahmad, 'DCT domain feature extraction scheme based on motor unit action potential of EMG signal for neuromuscular disease classification', Healthcare technology letters, 2014.

[16] K. A. I. Aboalayon, M. Faezipour, W. S. Almuhammadi, and S. Moslehpour, 'Sleep Stage Classification Using EEG Signal Analysis: A Comprehensive Survey and New Investigation', Entropy, 2016.

[17] H. Wu, Y. Qian, W. Zhang, and C. Tang, 'Feature extraction and identification in distributed optical-fiber vibration sensing system for oil pipeline safety monitoring', Photonic Sensors, 2017.

[18] V. K. Mishra, V. Bajaj, A. Kumar, D. Sharma, and G. K. Singh, 'An efficient method for analysis of EMG signals using improved empirical mode decomposition', Aeu-international Journal of Electronics and Communications, 2017.

[19] A. Sengur, Y. Akbulut, A. S. Ashour, Y. Guo, and V. Bajaj, 'Classification of amyotrophic lateral sclerosis disease based on convolutional neural network and reinforcement sample learning algorithm', 2017.

[20] Rohit Raja, Tilendra Shishir Sinha, Ravi Prakash Dubey (2015), Recognition of human-face from side-view using progressive switching pattern and soft-computing technique, Association for the Advancement of Modelling and Simulation Techniques in Enterprises, Advance B, Vol. 58, N 1, pp. 14-34, ISSN: -1240-4543.

[21] Tilendra Shishir Sinha, Raj kumar Patra, and Rohit Raja (2011) A Comprehensive analysis of human gait for abnormal foot recognition using Neuro-Genetic approach, International Journal of Tomography and Statistics (IJTS), Vol. 16, No. W11, pp. 56-73, ISSN: 2319-3339, http://ceser.res.in/ceserp/index.php/ijts.