ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 2, February 2016

A Review on Feature Extraction Methods

Kosha Patel

PG Student, Dept. of Biomedical Engineering, Government Engineering College, Gandhinagar, Gujarat, India[1]

ABSTRACT: The Electromyogram (EMG) signal is one of the bio signals utilized in helping humans to control equipment. For this we required to recognize the hand movement. In this direction the first step is feature extraction. The optimal feature is important for the achievement in EMG analysis and control. By this extracted feature we reduce the computational cost of a multifunction myoelectric control system. The goal of this paper is to define the methods and approaches which are most suited for extracting the features from EMG signal. The techniques discussed here are spectral approaches like Short Transform Fourier Transform (STFT) and wavelet transform analysis including Discrete Wavelet Packet Transform (DWPT) and Continuous Wavelet Transform (CWT).

KEYWORDS: EMG, Feature Extraction, Wavelet Transform, Fourier Transform

I. INTRODUCTION

EMG signals derive from neuromuscular activities of skeletal muscles; the information which is provided by EMG signal processing can be applied in a wide range of biomedical areas including rehabilitation, functional electrical simulation, clinical diagnosis, control of power prosthetic limbs and also fatigue analysis associated with muscle contraction [1]. EMG is a non-stationary signal due to variation in its frequency contents over time, considering this large amount of EMG signals, feeding the time sequence directly into a classifier is not practical. Hence The EMG signal is mapped into a smaller dimension vector or a feature vector prior to being classified. There exist large numbers of diversity methods for extraction of different features out of EMG signals of which five will be discussed through the presented review.

II. RELATED WORK

The success of any pattern classification system depends almost entirely on the choice of features used to represent the raw signals. It is desirable to use multiple feature parameters for EMG pattern classification since it is very difficult to extract a feature parameter which reflects the unique feature of the measured signals to a motion command perfectly. Considering the following feature parameters based on time and spectral statistics are chosen to represent the myoelectric pattern.[1]

Because of their computational simplicity, time domain features or linear techniques are the most popular in EMG pattern recognition. Integrated EMG, Mean absolute value, Mean absolute value slope, Variance of EMG, Root mean square, Waveform length, Zero crossing, Slope sign change, Willison amplitude, and Histogram of EMG are used to test the performance. All of them can be done in real-time and electronically and it is simple for implementation. [1] Features in this group are normally used for onset detection, muscle contraction and muscle activity detection. Moreover, features in frequency domain are used to represent the detect muscle fatigue and neural abnormalities, and sometime are used in EMG pattern recognition.

A. TIME DOMAIN FEATURE EXTRACTION

1) Integrated EMG

Integrated EMG (IEMG) is calculated as the summation of the absolute values of the sEMG signal amplitude. Generally, IEMG is used as an onset index to detect the muscle activity that used to oncoming the control command of assistive control device. It is related to the sEMG signal sequence firing point, which can be expressed as

$$IEMG = \sum_{n=1}^{N} |x_n|$$

where N denotes the length of the signal and x_n represents the sEMG signal in a segment.

2) *Mean Absolute Value*

Mean Absolute Value (MAV) is similar to average rectified value (ARV). It can be calculated using the moving average of full-wave rectified EMG. In other words, it is calculated by taking the average of the absolute value of sEMG signal. It is an easy way for detection of muscle contraction levels and it is a popular feature used in myoelectric control application. It is defined as

$$MAV = \frac{1}{N} \sum_{n=1}^{N} |x_n|$$

3) *Variance of EMG*

Variance of EMG (VAR) uses the power of the sEMG signal as a feature. Generally, the variance is the mean value of the square of the deviation of that variable. However, mean of EMG signal is close to zero. In consequence, variance of EMG can be calculated by

$$VAR = \frac{1}{N-1} \sum_{n=1}^{N} x_n^2$$

4) *Root Mean Square*

Root Mean Square (RMS) is modelled as amplitude modulated Gaussian random process whose RMS is related to the constant force and non-fatiguing contraction. It relates to standard deviation, which can be expressed as

$$RMS = \sqrt{\frac{1}{N} \sum_{n=1}^{N} x_n^2}$$

5) *Waveform Length*

Waveform length (WL) is the cumulative length of the waveform over the time segment. WL is related to the waveform amplitude, frequency and time. It is given by

$$WL = \sum_{n=1}^{N-1} |x_{n+1} - x_n|$$

6) *Zero Crossing*

Zero crossing (ZC) is the number of times that the amplitude value of sEMG signal crosses the zero y-axis. In EMG feature, the threshold condition is used to abstain from the background noise. This feature provides an approximate estimation of frequency domain properties. It can be formulated as

$$ZC = \sum_{n=1}^{N-1} [sgn(x_n \times x_{n+1}) \cap |x_n - x_{n+1}| \geq threshold$$

7) *Slope Sign Change*

Slope Sign Change (SSC) is similar to ZC. It is another method to represent the frequency information of sEMG signal. The number of changes between positive and negative slope among three consecutive segments are performed with the threshold function for avoiding the interference in sEMG signal. The calculation is defined as

$SSC = \sum_{n=2}^{N-1}[f[(x_n-x_{n-1})(x_n-x_{n+1})]]$

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 2, February 2016

8) Histogram of EMG

Histogram of EMG (HEMG) divides the elements in sEMG signal into *b* equally spaced segments and returns the number of elements in each segment. HEMG is an extension version of the ZC and WAMP features.

B. FREQUENCY DOMAIN FEATURE EXTRACTION

1) Autoregressive Coefficients

Autoregressive (AR) model described each sample of sEMG signal as a linear combination of previous samples plus a white noise error term. AR coefficients are used as features in EMG pattern recognition. The model is basically of the following form:

$$x_n = -\sum_{i=1}^{p} a_i x_{n-i} + w_n$$

Where $x_n$is a sample of the model signal, $a_i$is AR coefficients, $w_n$is white noise or error sequence, and *p* is the order of AR model.

2) Mean Frequency and Median Frequency

Traditional median frequency (MDF) and traditional mean frequency (MNF) are calculated based on power spectrum. We can calculate using the sEMG power spectrum *Pj*instead of amplitude spectrum *Aj*. They can be expressed as

$$\sum_{j=1}^{MDF} p_j = \sum_{j=MDF}^{M} p_j = \frac{1}{2}\sum_{j=1}^{M} p_j$$

$$MNF = \frac{\sum_{j=1}^{M} f_j p_j}{\sum_{j=1}^{M} p_j}$$

C. TIME-FREQUENCY REPRESENTATION

The fundamental purpose of feature extraction is to emphasize the important information in the measured signal while rejecting noise and irrelevant data. Time–frequency representation can localize the energy of the signal both in time and in frequency, thus allowing a more accurate description of the physical phenomenon. Time–frequency representation (TFR) generally requires a transformation that could be computationally heavy.

Among all different types of TFR, discrete, linear TFRs—short -time Fourier transform (STFT), wavelet transform (WT), and wavelet packet transform (WPT)—are preferable to quadratic TFRs, which are too computationally intense for real-time application. The fundamental difference between linear TFRs is in the manner in which they partition the time–frequency plane.

The STFT has a fixed tiling; once specified, each cell has an identical aspect ratio. The tiling of the wavelet transform is variable—the aspect ratio of the cells varies such that the frequency resolution is proportional to the center frequency. This tiling has been shown to be more appropriate for many physical signals, but the partition is nonetheless still fixed. The WPT provides an adaptive tiling—an over complete set of tiling are provided as alternatives, and the best for a given application is selected.

1) Short-time Fourier transform (STFT)

Most transforms, in their original form, assume that the signal under consideration is stationary. Generally this assumption fails in the case of EMG signal, except for short periods of time. Given a finite-length sequence *xi*, $i \in \{0, \ldots, L-1\}$, its discrete Fourier transform (DFT) is defined as

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

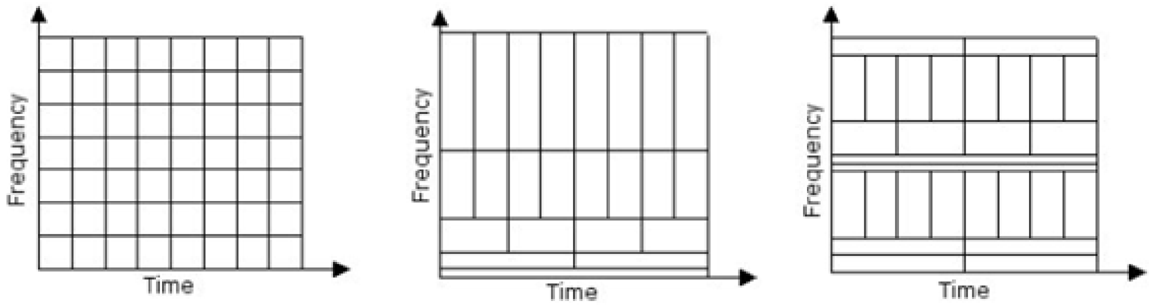
(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 2, February 2016

$$X[mF] = X[m] = \sum_{i=0}^{L-1} x[i]e^{-i2\pi(mF)(iT_s)}$$

where$L$ is the length of the sequence, $F = 1/LTs$ is the frequency sampling step size. The STFT consists of a series of DTFs, indexed with respect to Tsand F

$$STFT[k,m] = STFT[kT_s, mF] = \sum_{i=1}^{L-1} x[i]g[i-k]e^{-i2\pi mi/L}$$



(a) (b) (c)

Fig1 :Tiling of (a) STFT, (b) WT and (c) WP[2,3]

where$g$[i] is the window function.The temporal sampling step size is $T = K \cdot Ts$;if $K = 1$, the STFT is computed at every sample in time; if $K = L$, the successive analysis windows do not overlap. The resolution in time and frequency is lower bounded by the *time-bandwidth uncertainty principle* or *Heisenberg inequality*.

$$\Delta t.\Delta f \geq \frac{1}{4\pi}$$

An easy way to obtain a time-frequency representation is to slice the signal into non-overlapping adjacent blocks and expand each block independently as shown in figure 1.[2,3] For example, this can be done using a window function on the signal which is the indicator function of the interval [nT, (n+1)T], periodizing each windowed signal with period Tand applying an expansion such as the Fourier series on each periodized signal.The functions used in the expansion have changing time-frequency tiles because of the scaling. For small a (a <1), $\Psi_{a,b}(t)$ will be short and of high frequency, while for large a (a >1), $\Psi_{a,b}(t)$ will be long and of low frequency.

A Gaussian window allows a balanced time resolution and frequency resolution. The STFT has, among its other useful properties, a well-developed theory and can be computed very efficiently. The main constraint is that each cell in the time–frequency plane must have identical shape. In fact, as imposed by the temporal and frequency sampling steps, the time–frequency plane is divided into cells, each of which has a temporal width of T and a frequency height of F, and clearly the energy distribution of physical signals is not (in general) conveniently localized in region of fixed aspect ratio.

2) Wavelet transform (WT)

The WT overcomes the main drawback of the STFT by varying the time–frequency aspect ratio and by producing a good frequency resolution Δf in long time windows (low frequencies) and a good time localization Δtat high frequencies. This produces a tiling of the time–frequency plane that is appropriate for most physical signals.[2,3,4]

The continuous wavelet transform is defined as

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 2, February 2016

$$CWT_x(\tau,\alpha) = \frac{1}{\sqrt{\alpha}}\int x(t)\Psi(\frac{t-\tau}{\alpha})$$

where$\Psi(t)$ is the *mother wavelet*, which has the property that the set $\{\Psi a,b\}$ forms an orthonormal basis in $L^2(\Re)$ $(a,b\in\Re,\ a\neq 0)$.

3) *Wavelet Packet Transform (WPT)*

WPT is a generalized version of the CWT and the DWT. Because the CWT is redundant, the tiling of the time–frequency plane is configurable. The basis for the WPT is chosen using an entropy-based cost function. While the WT is extremely computationally efficient (it takes less than 2 ms to calculate the WT of a 256-ms stream of data), the WPT demands substantially greater computation, about 200 ms on a record length of 256 ms. Both WT and WPT have been tested on EMG signals[5,6], showing that time–frequency representation, together with a linear dimensionality reduction, is capable of a better description of the intended movement.

III. CONCLUSION

The extraction of accurate features from the EMG signals is the main kernel of classification systems and is essential to the motion command identification. But the non-stationarity of the EMG signal makes it difficult to extract feature parameters precisely with the block processing stationary model such as an autoregressive (AR) model. And it is very difficult for one feature parameter to reflect the unique feature of the measured EMG signals to a motion command perfectly. Once a feature set has been chosen, a suitable pattern classifier can be used to determine class output. CWT was the most satisfactory analysis tool for this purpose as it produced a better display of wavelet transform. It gave the most detailed information in terms of range of frequencies for each scale used. The EMG feature like Root mean square, Median frequency, Mean frequency are extracted with high accuracy using Continuous Wavelet Transform.

REFERENCES

1. Phinyomark, Angkoon, ChusakLimsakul, and PornchaiPhukpattaranont. "A novel feature extraction for robust EMG pattern recognition." *arXiv preprint article no:0912.3973* (2009).
2. Zecca, Micera, et al. "Control of multifunctional prosthetic hands by processing the electromyographic signal." *Critical Reviews™ in Biomedical Engineering* Vol.30 no..4 (2002).
3. Vetterli, Martin, and JelenaKovacevic. *Wavelets and subband coding*. No. LCAV-BOOK-1995-001. Prentice-hall, 1995.
4. Kilby, J., and K. Prasad. "Continuous Wavelet Analysis and Classification of Surface Electromyography Signals." *International Journal of Computer and Electrical Engineering Vol.*5 no.1 (2013)
5. Englehart K, Hudgins B, Parker PA. A wavelet-based continuosclassifi cation scheme for multifunction myoelectric control. IEEE Trans Biomed Eng 2001; article no. 48 pg:302–311.
6. P. Parker, K. Englehart, and B. Hudgins, “Myoelectric Signal Processing for Control of Powered Limb Prostheses,” *Journal of Electromyography and Kinesiology*, vol. 16, no. 6, pp. 541-548, Dec 2006.