

Hilbert-Huang transform based physiological signals analysis for emotion recognition

Cong Zong and Mohamed Chetouani

Université Pierre et Marie Curie

Institut des Systèmes Intelligents et de Robotique, CNRS UMR 7222

4 Place Jussieu 75252 Paris Cedex 05, France

cong.zong@isir.fr, mohamed.chetouani@upmc.fr

Abstract—This paper presents a feature extraction technique based on the Hilbert-Huang Transform (HHT) method for emotion recognition from physiological signals. Four kinds of physiological signals were used for analysis: electrocardiogram (ECG), electromyogram (EMG), skin conductivity (SC) and respiration changes (RSP). Each signal is decomposed into a finite set of AM-FM mono components (fission process) by the Empirical Mode Decomposition (EMD) which is the key part of the HHT. The information components of interest are then combined to create feature vectors (fusion process) for the next classification stage. In addition, classification is performed by using Support Vector Machines (SVM). The classification scores show that HHT based methods outperform traditional statistical techniques and provide a promising framework for both analysis and recognition of physiological signals in emotion recognition.

Keywords—Hilbert-Huang transform, physiological signals, emotion recognition.

I. INTRODUCTION

In human-machine interaction, emotion recognition is now recognized as a fundamental element. A possible role of emotion is to make the interaction between humans and intelligent machines (agents, robots) more natural and sociable [1]. In other words, these machines should be able to interact naturally with users in order to improve the services offered. This process requires advanced signal analysis and pattern recognition techniques in order to detect and interpret users' emotional states. There are many ways that humans display their emotions. Physiological responses are natural responses that are privileged relative to other channels of communication. However, as mentioned in [2], biosignals have received less attention than verbal and non-verbal communication channels. One of the reasons can be attributed to the heterogeneity of physiological signals. In this paper, we propose to combine advanced signal processing and machine learning techniques for the characterization of physiological signals.

Feature extraction is a key step of pattern recognition. If the features extracted are carefully chosen, it is expected that the features set will extract relevant information from the input data in order to perform the desired task. There are many techniques proposed to extract features from physiological signals in the literature [2], [3], [4], [5]. Kim et al. [2] extracted different features: conventional statistics, spectrum, entropy from various analysis domains (time series, frequency domain). A total of 110 features were obtained from four physiological signals. After the computation of a large number of features, the best emotion-relevant features

were selected and then used for classification. Maaoui et al. [4] acquired six features in time domain for each signal from 10 subjects in six different affective states. Two pattern recognition methods have been tested: SVM and Fisher linear discriminant based classifiers. Although the state-of-the-art systems seem to perform well enough for emotion recognition, they have disadvantages. There are some crucial restrictions of the Fourier transform: the system must be linear; and the data must be strictly stationary; otherwise, the features extracted via Fourier transform do not have any physiological significance [6]. Physiological signals are nonlinear and non-stationary signals, for which Fourier transform is unsuitable. So, the conventional methods may not provide accurate frequency resolution [7]. It is desirable to develop quantitative methods which do not assume stationarity.

Huang et al. [6] have proposed the Hilbert-Huang transform method (HHT) as a new tool for the analysis of nonlinear and non-stationary data. Unlike the Fourier transform, which is predicated on a priori selection of basis functions that are either of infinite length or have fixed finite widths, Empirical Mode Decomposition (EMD) decomposes a signal into finite basis functions called the intrinsic mode functions (fission process). After the fission process, the intrinsic mode functions (IMFs) of interest are then combined in an ad hoc or automated fashion in order to provide greater knowledge about a process in hand (fusion process) [8]. Recently HHT has been applied to process ECG and EMG [7], [9], [10], [11]. Xie et al. [7] used HHT to estimate the mean frequency and it has been applied to fatigue EMG signal analysis. The method derived via HHT showed low variability in terms of robustness to the length of the analysis window. The mean frequency of physiological signals seem to provide interesting informations and are also investigated in this paper.

Section II presents the data corpus of Augsburg [3] and the four physiological signals used in this work. Section III describes feature extraction techniques based on the Hilbert-Huang transform (HHT) method and others derived from the state-of-the-art (baseline). The experimental protocols and results are discussed in section V and section VI gives conclusions and some directions for our future works.

II. DATA CORPUS

We use the data of University of Augsburg [3]. In this database, the physiological signals were recorded at the same

time when the subject was listening to different musics. Four music songs were used to induce him to feel four different emotions: joy, anger, sadness and pleasure (see Fig.1). After having listened the music, he himself carefully annotated the physiological signals on four categories (four emotions). There are 25 recordings for each emotion. Moreover, the raw signals were trimmed to a fixed length of two minutes.

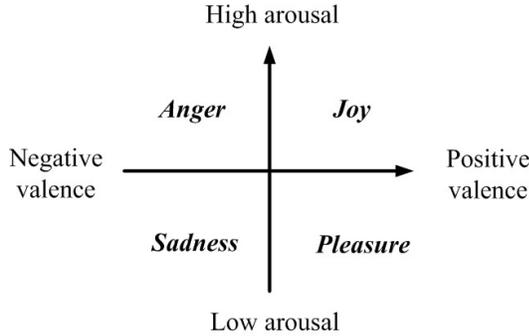


Fig. 1. Emotion models

The following four physiological signals are used :

- **Electrocardiogram (ECG):** An electrocardiogram is a simple test that records the electrical activity of the heart. The electrical activity is related to the impulses that travel through the heart and determine the heart's rate and rhythm. It is able to give a precise estimate of heart rate, R-R interval, and heart rate variability. The changes in heart rate have been used as indicators of different emotions.
- **Electromyogram (EMG):** Electromyography is a technique for evaluating and recording the electrical potential generated by muscle cells. In psychophysiology, EMG is often used to find the correlation between cognitive emotion and physiological reactions [12].
- **Skin conductivity (SC):** The skin conductivity response is one of the most robust and well studied physiological response. It measures the electrical resistance of the skin. The magnitude of this electrical resistance is effected not only by the subject's general mood, but also by immediate emotional reactions. Thus, it can be used as emotional indicator [13].
- **Respiration changes (RSP):** The amount of stretch in the elastic is measured as a voltage change and recorded. The depth the subject's breath and subject's rate of respiration are the most common measures of RSP. Moreover, other measures can be affected by RSP, for example, ECG measurements, because RSP is closely linked to cardiac function [2].

III. FEATURE EXTRACTION

Before feature extraction, a pre-processing step is required in order to reduce noises and artifacts. At first, we used

adaptive low-pass filters to remove the peaks (noises) of signals. For the ECG signal, an R-peak detector was used to extract the R-R intervals. The sampling period of the R-R interval signals is by definition irregular and HHT provides a suitable framework for the analysis of such signals. In this paper we re-sampled the signal in order to compare the proposed approach with existing techniques. We employed the Berger's algorithm [14] with resampling frequency of 4 Hz. For the SC signal, we subtracted the baseline to consider only relative amplitudes. Moreover, the EMG signal is normalized by its mean and standard deviation.

A. State-of-the-art approach (Baseline)

Even if various features have been employed in [2], [3], we focus on the statistical features in the time domain (max, min, mean, median, standard deviation and range) since they are the most commonly used features. However, for some signals such as ECG and RSP, frequency based features have been extracted such subbands analysis similarly to the methods presented in [2], [3]. The classification results show that the best domain depends on the signal used ECG: 45% (frequential), 59% (time domain), RSP: 62%, 42%. The state-of-the-art architecture is presented in Fig. 2(a).

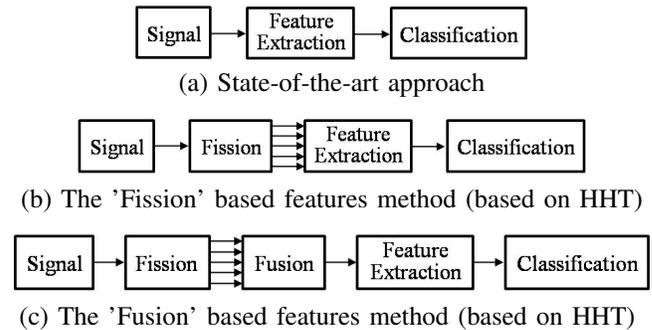


Fig. 2. Three classification systems

B. Hilbert-Huang transform

In this section, we propose a novel method based on the Hilbert-Huang transform (HHT) [6] to analyze physiological signals. The HHT comprises the 'Empirical Mode Decomposition' (EMD) and Hilbert transform. The key part of the HHT is the EMD method with which any data set can be decomposed into a finite and often small number of 'Intrinsic Mode Functions' (IMF) (fission process). Each IMF satisfies two conditions [6] : (1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The EMD provides a framework for both information fission and fusion [8]. For the fission approach, we extract features from each IMF and the feature vector is composed by the

combination of these features. The fusion approach aims at computing directly a feature vector by merging information from the IMFs. The proposed methods shown in Fig.2 are presented in the following sub-sections.

1) *Fission approach*: Given a signal $x(t)$, the effective algorithm of EMD is as follows :

- Step 1 Identify all extrema (maxima and minima) of $d_0(t) = x(t)$.
- Step 2 Interpolate between the maxima and connect them by a cubic spline curve. The same applies for the minima in order to obtain the upper and lower envelopes $e_u(t)$ and $e_l(t)$, respectively.
- Step 3 Compute the average $m(t) = \frac{e_u(t)+e_l(t)}{2}$.
- Step 4 Extract the detail $d_1(t) = d_0(t) - m(t)$.
- Step 5 Iterate steps 1-4 on the residual until the detail signal $d_k(t)$ satisfies the above definition of an IMF: $c_1(t) = d_k(t)$. The IMF requirements are checked indirectly by evaluating a stopping criteria. (Sifting process)
- Step 6 All above steps are iterated until the final residual $r_n(t) = x(t) - c_n(t)$ is a monotonic function. The last residual is considered as the trend.

The stopping criteria is important for estimation of IMFs, which can influence number of IMFs components. Rilling et al. [15] proposed a criterion based on two thresholds θ_1 and θ_2 which guarantee globally small fluctuations in the mean while taking into account local large excursions. The mode amplitude $a(t) = \frac{e_u(t)-e_l(t)}{2}$ is introduced so the evaluation function is $\sigma(t) = |\frac{m(t)}{a(t)}|$. The sifting process is iterated until $\sigma(t) < \theta_1$ for some prescribed fraction $(1 - \alpha)$ of the total duration, while $\sigma(t) < \theta_2$ for remaining fraction. The default values are set to $\alpha = 0.05$, $\theta_1 = 0.05$ and $\theta_2 = 10\theta_1$.

Based on the above algorithm, the original $x(t)$ can be exactly reconstructed by a linear superposition:

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (1)$$

where n is the number of IMFs.

The second step of the HHT is the Hilbert transform which is applied to each IMF components:

$$H[c_i(t)] = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{c_i(t')}{t-t'} dt' \quad (2)$$

PV denotes the Cauchy principal value of this integral. An analytic signal is defined as:

$$z_i(t) = c_i(t) + jH[c_i(t)] = a_i(t)e^{j\theta_i(t)} \quad (3)$$

To extract instantaneous frequency information for each IMF, we calculate the derivative of the phase:

$$f_i(t) = \frac{1}{2\pi} \frac{d\theta_i(t)}{dt} \quad (4)$$

The fission approach is based on the decomposition of the signal into n IMFs by applying EMD. The number of IMFs

for each biosignal varied from n_{min} to n_{max} (n_{min} and n_{max} are respectively minimum and maximum number of IMFs for each biosignal). According to Wang et al. [16], we calculated root mean square, maximum amplitude of analytic signals, mean instantaneous frequency and weighted mean instantaneous frequency of each IMF. Four features were extracted from each IMF. The feature vectors were then constructed by the features of the first $(1, \dots, n_{min})$ IMFs. Finally, by comparing performances of features based on different number $(1, \dots, n_{min})$ of IMFs, we chose the features of the first m IMFs which yielded best performance for each biosignal (m is the optimal number of IMFs). We show the block diagram of the classification system in Fig.2(b).

2) *Fusion approach*: The fusion process aims at merging information from the different IMFs and it is computed by the characterization of the mean frequency (MNF) of a given signal. The classification system shown in Fig.2(c) is based on the following steps: fission (HHT), fusion (MNF estimation), feature extraction (statistics: max, min, mean, median, standard deviation and range), and classification. The main difference with the fission approach is the estimation of the MNF, which requires the following steps. Each signal is first segmented by using half-overlapped windows. EMD is then applied to each segment resulting on several IMFs (fission process). Then the weighted mean instantaneous frequency $WMIF(i)$ of each IMF ($c_i(t)$) with N samples is defined by :

$$WMIF(i) = \frac{\sum_{j=1}^N f_i(j)a_i^2(j)}{\sum_{j=1}^N a_i^2(j)} \quad (5)$$

Xie et al. [7] showed that $WMIF(i)$ can be used as a measure of the mean frequency (MNF) of the signal in the frequency band (fusion process), so the MNF with n IMFs is defined as:

$$MNF = \frac{\sum_{i=1}^n \|a_i\| WMIF(i)}{\sum_{i=1}^n \|a_i\|} \quad (6)$$

IV. CLASSIFICATION METHOD

There are many classification methods that can be employed [17]. In this study, the Support Vector Machine (SVM) classifier is used [18]. In this paper, we select linear kernel which performs well on our database. The discriminant function of SVM is given in following:

$$f(x) = \text{sgn} \left(\sum_{j=1}^M y_j \alpha_j K(\vec{x}_j, \vec{x}) + b \right) \quad (7)$$

where M is the number of training examples, x_j is the j_{th} training example, and y_j is the correct output of the SVM for the i_{th} training example. K is a kernel function that is used to transform data, α_j are Lagrange multipliers of a dual optimization problem.

In our database, records will be classified into four classes. Therefore, we used one-against-one decomposition to train the multi-class SVM classifier [19]. The classification into

N_class classes is decomposed into $N_rule = (N_class - 1) * N_class / 2$ binary problems. In addition, a 10-fold cross validation technique was used to estimate the classification scores. The feature vectors are normalized (zero mean, unit standard deviation).

V. EXPERIMENTAL RESULTS

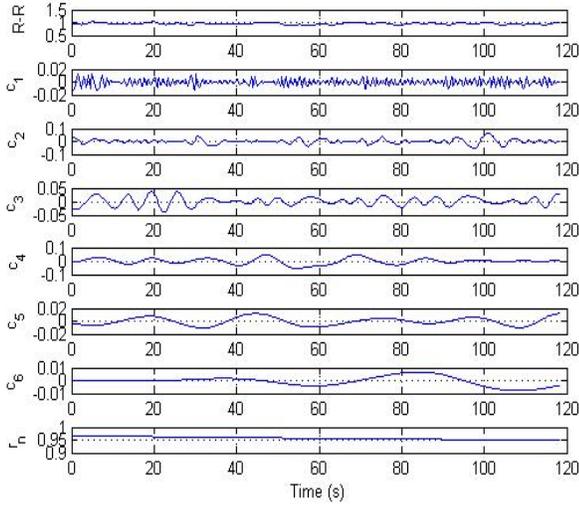


Fig. 3. The decomposition of R-R interval signal (emotion of sadness)

An example of R-R interval signal and its IMFs by using EMD is shown in Figure 3. We can observe that EMD performs a fission process : the R-R interval signal is decomposed into six IMFs components and a residual component. The instantaneous frequency of each IMF is then obtained by the Hilbert transform and Figure 4 shows the results for the first three IMFs.

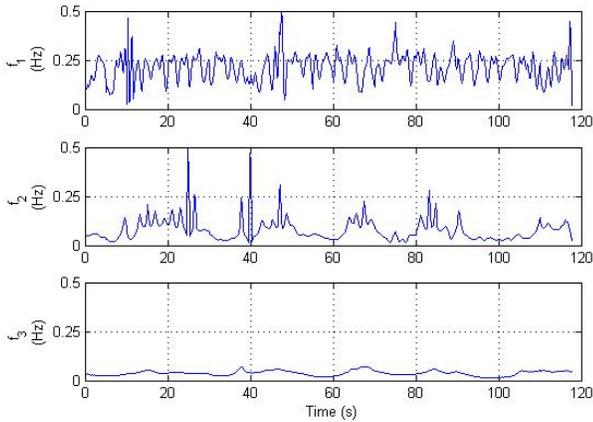


Fig. 4. Instantaneous frequency (normalized) for the first three IMFs for R-R interval signal

	IMF_1 (4 features)	$IMF_{1,2}$ (8 features)	$IMF_{1,2,3}$ (12 features)	$IMF_{1,2,3,4}$ (16 features)
ECG (R-R)	57%	69%	69%	63%
SC	35%	34%	39%	-
EMG	52%	49%	52%	49%
RSP	60%	57%	52%	-

Table 1. Comparison of classification results of different number of IMFs based features ('Fission' based features)

As mentioned before (cf. III-B.1), we classified four emotions by different number of IMFs ($1, \dots, n_{min}$) estimated by HHT of each biosignal and compared their classification results which are shown in Table 1. The minimum number of IMFs (n_{min}) of ECG signal and EMG signal is four; three for SC and RSP signals. After comparison, the number of IMFs was optimized by the best performance of each biosignal (shown in bold). It can be seen that the information of the first two or three IMFs provides the best performance (69%) for ECG signal; the first three for SC signal; the first one or three for EMG signal and the first one or three for RSP signal. Finally, if classification results of two sets of features for same biosignal were same, the configuration providing the lowest number of features is chosen.

	Baseline	HHT-based	
		'Fission' based features (n: Optimal number of IMFs)	'Fusion' based features MNF (Optimized window)
ECG (R-R)	59% (TD) 45% (FD)	69% (2)	56% (64 samples)
SC	42% (TD)	39% (3)	30% (1024 samples)
EMG	39% (TD)	52% (1)	44% (1024 samples)
RSP	42% (TD) 62% (FD)	60% (1)	59% (512 samples)

Table 2. Recognition rates of four emotions for four biosignals by using SVM (TD : Time Domain ; FD : Frequency Domain)

Table 2 gives the percentage of correctly classified examples, respectively, for four signals by different feature extraction techniques mentioned in section III including also the state-of-the-art methods in both time and frequency domains. For the fusion based HHT features method (MNF), we optimized the length of the analysis window resulting in an adaptive window size for the segmentation of each signal. As a result, the accuracy rates of this method are better than the baseline for EMG. Concerning the fission approach,

we focused on information of the first m IMFs (m is the optimal number of IMFs). This approach provides the best scores regarding the other methods (state-of-the-art and fusion approaches). For instance, the fission approach outperforms both time and frequential techniques for the ECG signal characterization showing the interest of the time-frequency description defined in the HHT. However, this combination requires more investigations since the results are dependent on the signals: for SC, best results are obtained by traditional methods. One of the reasons might be the less rhythmic behaviors of this signal compared to the other ones (R-R, EMG and RSP).

We then combined all features of four biosignals of each method (the state-of-the-art method, the fission and the fusion based HHT features method), and also classified by using the SVM. The classification results are given in Table 3.

	Baseline (32 features)	HHT-based	
		'Fission' based features (28 features)	'Fusion' based features (24 features)
Fusion of four biosignals	71%	76%	62%

Table 3. Classification results of feature fusion (four biosignals)

It can be seen, that the highest recognition accuracy is obtained by using the fission based HHT features method with 28 features, 76% of test cases are correctly classified. Although the state-of-the-art method achieves a recognition rate of 71%, it employs more features than the proposed methods based on fusion and fission approaches.

The results of tables 2 and 3 show that the fission process obtained by HHT can be exploited for pattern recognition problems such as emotion recognition from heterogeneous physiological signals. However, the main problem relies on the exploitation of the features extracted from this process. And the results show that efficient performances can be achieved by an adequate selection of IMFs and by the early and simple fusion of the feature vectors extracted from each IMF.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we analyzed physiological signals to recognize four different emotions. Two methods based on the Hilbert-Huang transform (HHT) were proposed based on the properties of this transformation: fission and fusion approaches. The former is based on the selection of IMFs and early fusion of feature vectors. The latter is based on the estimation of the mean frequency of signal. The approaches have been compared to state-of-the-art techniques. The experimental results show that HHT features outperform traditional methods. Additionally, the proposed feature extraction scheme allows to analyze the heterogeneous signals in a similar framework based on the estimation of instantaneous amplitudes and frequencies. Our future works are devoted to the investigation of combination

methods for exploiting the fission-fusion processes provided by HHT.

ACKNOWLEDGMENTS

We would like to thank Johannes Wagner from Institute of Computer Science of University of Augsburg for sharing the data set and the results of their work. We also acknowledge our colleagues in the MIRAS project (Multimodal Interactive Robot of Assistance in Strolling) for fruitful discussions. This work has been supported by French National Research Agency (ANR) through TecSan program (project MIRAS n° ANR-08-TECS-009).

REFERENCES

- [1] H. R. Kim, K. W. Lee, and D. S. Kwon, "Emotional interaction model for a service robot," in *Proceedings of the 14th IEEE International Workshop on Robot and Human Interactive Communication, Nashville, Tennessee, USA, Aug 2005*, pp. 672–678.
- [2] J. Kim and E. André, "Emotion recognition based on physiological changes in listening music," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2067–2083, December 2008.
- [3] J. Wagner, J. Kim, and E. André, "From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification," *IEEE International Conference on Multimedia & Expo (ICME 2005)*, pp. 940–943, 2005.
- [4] C. Maaoui, A. Pruski, and F. Abdat, "Emotion recognition for human-machine communication," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1210–1215, 2008.
- [5] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1175–1191, 2001.
- [6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, March 1998, pp. 903–995.
- [7] H. Xie and Z. Wang, "Mean frequency derived via hilbert-huang transform with application to fatigue emg signal analysis," *Computer Methods and Programs in Biomedicine*, vol. In Press, Corrected Proof.
- [8] D. Mandic, M. Golz, A. Kuh, D. Obradovic, and T. Tanaka, "Complex empirical mode decomposition for multichannel information fusion," *Signal Processing Techniques for Knowledge Extraction and Information Fusion*, pp. 243–260, 2008.
- [9] J. C. Echeverría, J. A. Crowe, M. S. Woolfson, and B. R. Hayes-Gill, "Application of empirical mode decomposition to heart rate variability analysis," *Medical and Biological Engineering and Computing*, pp. 471–479, 2001.
- [10] E. P. S. Neto, M. A. Custaud, J. C. Cejka, P. Abry, J. Frutoso, C. Gharib, and P. Flandrin, "Assessment of cardiovascular autonomic control by the empirical mode decomposition," *Methods of information in medicine*, pp. 60–65, 2004.
- [11] H. Li, L. Yang, and D. Huang, "Application of hilbert-huang transform to heart rate variability analysis," *The 2nd International Conference on Bioinformatics and Biomedical Engineering*, pp. 648–651, 2008.
- [12] D. M. Sloan, "Emotion regulation in action: emotional reactivity in experiential avoidance," *Behaviour Research and Therapy*, vol. 42, pp. 1257–1270, 2004.
- [13] P. J. Lang, "The emotion probe. studies of motivation and attention," *The American psychologist*, vol. 50, no. 5, pp. 372–385, May 1995.
- [14] R. D. Berger, J. P. Saul, and R. J. Cohen, "An efficient algorithm for spectra analysis of heart rate variability," *IEEE Trans. Biomed. Eng.*, vol. 33, no. 9, pp. 900–904, 1986.
- [15] G. Rilling, P. Flandrin, and P. Gonçalvès, "On empirical mode decomposition and its algorithms," in *Proceedings of the 6th IEEE/EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP '03), Grado, Italy, 2003*.

- [16] N. Wang, Ambikairajah, E. Celler, B. Lovell, and N.H., "Accelerometry based classification of gait patterns using empirical mode decomposition," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 617–620, 2008.
- [17] A. Cornuéjols, L. Miclet, and Y. Kodratoff, *Apprentissage Artificiel: Concepts et algorithmes*. Eyrolles, August 30, 2002.
- [18] C. J. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, June 1998.
- [19] R. Debnath, N. Takahide, and H. Takahashi, "A decision based one-against-one method for multi-class support vector machine," *Pattern Anal. Appl.*, vol. 7, no. 2, pp. 164–175, 2004.