Technology and Health Care 24 (2016) S551–S559
DOI 10.3233/THC-161181
IOS Press

Image classification based on ICA-WP feature of EEG signal

Wei Zhu*, Han Zhang, Weiping Ni, Xiong Xu and Junzheng Wu
Northwest Institute of Nuclear Technology, Xi'an, Shaanxi, China

Abstract. In this paper, a method for classifying electroencephalographic (EEG) recordings with images as stimulation is introduced, which aims at selecting the target images. EEG recordings to be processed are referred to the onset of the test images with a single stimulation so as to avoid spending extra time on repeating images. Independent component analysis (ICA) is used to reduce the redundancy of EEG recordings, and wavelet packet (WP) analysis is efficient for dealing with the non-stationary character of brain activity. Feature vectors are extracted by a method that combines these two algorithms. The support vector machine is used as a classifier, carrying out the classification result. The experimental results demonstrate that the accuracy of this method's image classification is affected very little by different classifier parameters. The best result achieves 90% accuracy, which indicts it is feasible for classifying images with a single stimulation.

Keywords: EEG, independent component analysis, wavelet-packet, support vector machine

1. Introduction

As technology continuously improves, brain-signal acquisitions have become more convenient, noninvasive, and inexpensive. Furthermore, there are current methods available that make it easier to research brain activity through scalp electroencephalography (EEG) [1]. Thus, EEG is garnering a greater amount of attention for improving people's performances by engaging brain activities.

EEG offers high time precision, despite a coarse spatial resolution, and it is the main method used in current studies. At the present, there have been numerous applications that aid in improving communications and activities for disabled people, such as the brain-computer interface "P300 speller" [2] that is used for letter inputting, wheelchairs that are controlled by brain signals [3], and even interactive games [4]. In addition, application could be developed that could help to improve the performance of regular people with efficiency and accuracy. In the domain of image estimation, people are required to identify targets from a large amount of images, which demands extensive concentration, so people easily become tired. Furthermore, the image content is so varied that it is difficult for computer algorithms, which are highly efficient in many applications, to recognize the targets with a high accuracy. There is a need to involve brain research for assistance.

Most well-known applications are based on the superposition principle to improve the signal-to-noise-ratio of event-related potentials (ERPs) [5–8]. Contrary to those applications, image classifying requires a high efficiency because of the large amount of samples. Therefore, repeating a stimulation is not feasible in this field; rather, a real-time method for feature extraction and classifier is needed.

*Corresponding author: Wei Zhu, Northwest Institute of Nuclear Technology, Xi'an 710024, Shaanxi, China. Tel.: +86 02984767440; E-mail: zhuwei@nint.ac.cn.

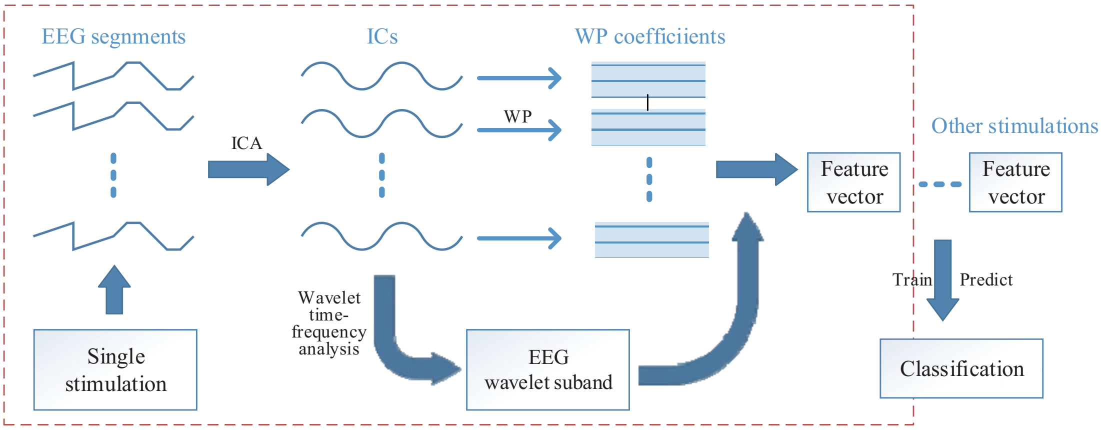


Fig. 1. Flowchart detailing the relationships between methods.

A variety of feature extraction methods are used: time-domain, frequency-domain, and time-frequency methods. Brain activities are full of spontaneous potentials whose amplitudes are much higher than those of evoked potentials. Since the spontaneous potentials are non-stationary, it is difficult to extract stable characters with the pure time-domain methods or to reflect changes of the brain signal along time with the pure frequency-domain. However, the time-frequency method is advantageous when dealing these issues. The EEG recordings have a large amount of data, constructed by electrical potentials in many different locations on the scalp. These potentials are presumably generated by mixing some underlying components of brain activity as well as artifact disruptions. The independent component analysis (ICA) [9] algorithm, noted for blind source separation, is used to make an abstract of signals in order to reveal interesting information and remove the artificial components.

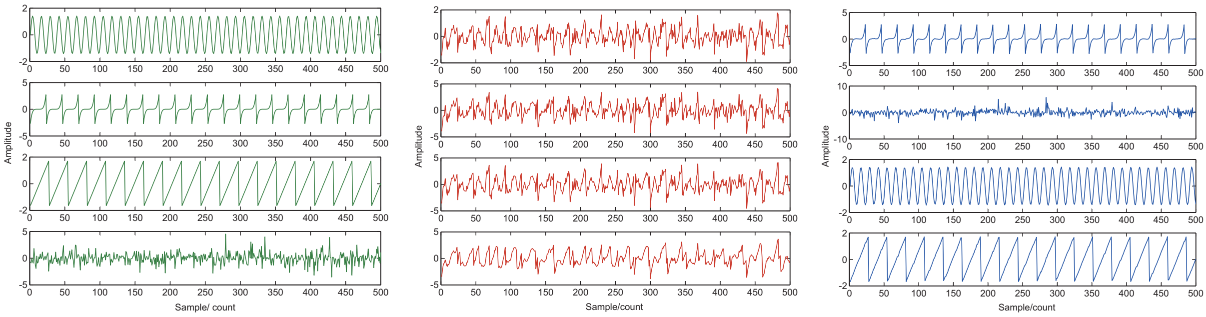
There are many classification algorithms, linear or not, such as the artificial neural network, k-means, and Bayesian algorithms. The expertise of the support vector machine (SVM) [10] is non-linear and multi-dimensional classification problems, and thus, over-learning or minima problems will not be an issue.

In this paper, an EEG recording-based target image classifying algorithm that uses a new type of feature extraction that combines wavelet packet (WP) [11] and ICA is introduced. A discrete wavelet decomposition is first applied in order to eliminate a high frequency component, known as detail coefficients, of the EEG signal. Wavelet packet decomposition [11] is implemented on independent components (ICs), which are the output of ICA on the EEG array after the artificial ICs are removed. Then, the WP coefficients are passed to the SVM, which is used as the classifier.

The background and the content of this paper is presented in the first section. A general view of the algorithms of feature extraction and classification are introduced in the second section. The experimental results and the discussion are presented in the third section. Finally, the conclusion is given at the end of this paper.

2. Methods

Figure 1 diagrams the relationships between the methods used in this paper. The EEG segments in one stimulation event generate one feature vector. The feature vectors are then divided into two sets for training and predicting. The details are explained.



(a) independent signal sources (b) mixed signal records as input (c) components acquired from ICA

Fig. 2. Demonstration of ICA with simulated signals.

2.1. Data acquisition

The EEG dataset used in this paper comes from an experiment conducted by the Swartz Center for Computational Neuroscience of the Institute for Neural Computation, the University of California, San Diego [12].

In their experiment, small image clips from a publicly available satellite image of London that had superimposed small target airplane images on it were shown to participants, which is referred to as the oddball paradigm. EEG was collected using a BIOSEMI Active View 2 system with 256 electrodes mounted in a whole-head elastic electrode cap (E-Cap, Inc) with a custom near-uniform montage across the scalp, neck, and bony parts of the upper face. Data acquisition was performed by USB using a customized acquisition driver at a 256 Hz sampling rate with 24-bit digitization.

Details of the data reorganizing and processing in the experiment are depicted at the beginning of Section 3.

2.2. Feature extraction

2.2.1. ICA

ICA [9] is an algorithm that solves the blind source separation problem, which attempts to decompose a multivariate signal into additive independent components (ICs). The ICA is based on the assumptions that the source signals are non-gaussian and are statistically independent from each other.

The ICA extracts the independent components in such a way that the statistical independence of the estimated components is maximized. The definition of independence may be chosen, and it is the choice of definition that determines the form of the ICA algorithm. There are two broad definitions of independence for ICA: the minimization of mutual information and the maximization of non-gaussianity.

The observed data $X(t) = [X_1(t), \cdots, X_M(t)]^T$ can be transformed into independent components $S(t) = [S_1(t), \cdots, S_N(t)]^T$. The generative formula can be written as:

$$X(t) = AS(t), M \geqslant N \tag{1}$$

A is the mixing matrix. The observed data can be represented by the mixing vector.

The recordings are presumable mixtures of some underlying components of brain activities, such as spontaneous and evoked potentials. Although the goal is to find the original brain activities, only a mixture of the components can be observed. ICA can reveal interesting information about the brain activity by accessing its independent components, as shown in Fig. 2.

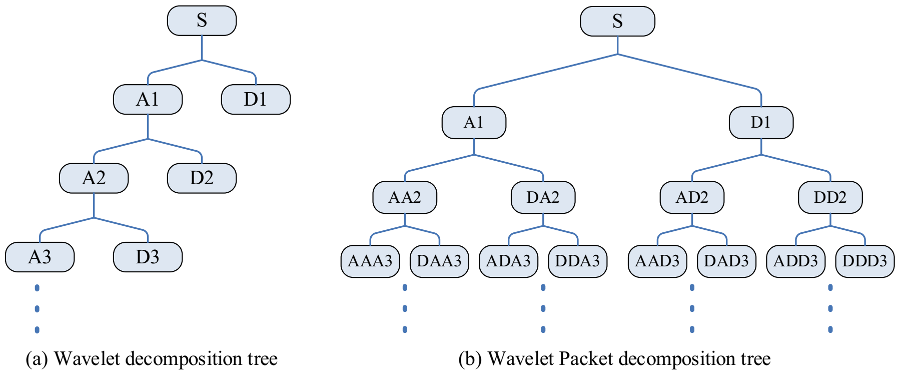


Fig. 3. Schematic diagram of Wavelet and Wavelet Packet decomposition tree ('A' represents approximation, and 'D' represents detail.)

A major problem in the analysis of EEG recordings is that the activities caused by artifacts, such as eye movements or sensor movement from sweating, typically have much higher amplitudes than those generated by neural sources. To identify artificial ICs, a completely automatic algorithm based on ICA (ADJUST [13]) is used to identify artificial ICs by combining stereotyped artifact-specific spatial and temporal features. Features are optimized to capture blinks, eye movements, and generic discontinuities. Once the artificial ICs are identified, they can be simply removed from the data while leaving the activity caused by neural sources nearly unaffected.

2.2.2. Time-frequency analysis

In the last decades, wavelets analysis [11] has become very popular. Different from Fourier transform, which separates the whole signal into pure harmonic waves, the main characteristic of wavelets is the possibility of providing a multi-resolution analysis of signals.

In numerical analysis and functional analysis, a discrete wavelet transform (DWT) is any wavelet transform where the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution, which captures both the frequency and time information. A family of wavelets is an orthogonal basis. By combining the wavelets, signals can be represented.

However, wavelets are ill suited to represent oscillatory patterns. In DWT, each level is only calculated with the previous wavelet approximation coefficients through discrete-time low and high pass quadrature mirror filters. Oscillatory variations of intensity can only be described by small-scale wavelet coefficients, which are represented as detail coefficients and will not be calculated at the next level, but those small-scale coefficients carry very little energy even at high sampling rates.

Wavelet Packet Decomposition (WPD) [11], sometimes known as just Wavelet Packets (WP) or Subband Tree, is a generalization of the classical wavelet decomposition that offers a richer signal analysis. Based on wavelet transform theory, the discrete-time signal is passed through more filters than the DWT. As shown in Fig. 3, both the detail and approximation coefficients are decomposed to create the full binary tree. In that case, the details as well as the approximations can be split. The WPD function can be written as shown in Fig. 3.

2.2.3. ICA-WP feature

Brain activity is highly complex and non-stationary; thus, it is unrealistic to extract features out of the EEG signals with a single method. Since the two aforementioned methods both have their own specialties, by combining ICA and WP, the extracted feature will be stable and robust.

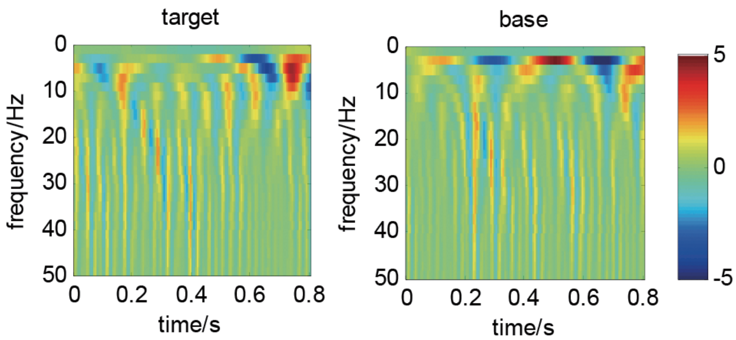


Fig. 4. Wavelet time-frequency analysis comparison between the target event and base event in IC No.1. (Each IC has a particular character; the target event contains no salient difference from the base in the high frequency area.)

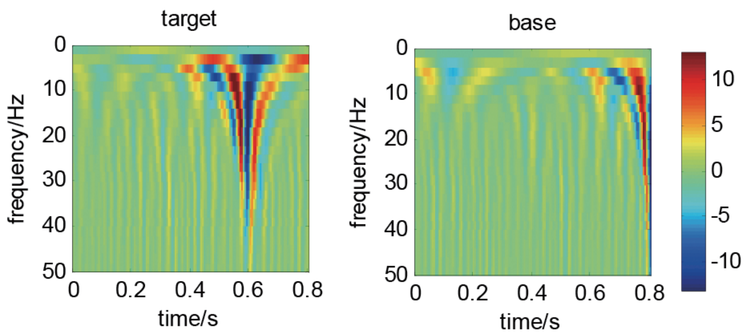


Fig. 5. Wavelet time-frequency analysis comparison between the target event and base event in IC No.8. (Each IC has a particular character; the target event contains no salient difference from the base in the high frequency area.)

The ICs that construct the recording array were decomposed after the ICA procedure. As shown in Figs 4 and 5, the time-frequency figures of different ICs have a significant specialty of their own with CWT processing. Each IC points to a particular character of the origin signals; thus, ICA is used to reduce the data redundancy among the EEG channels and separate noise components.

Since not all ICs are event-related components and many are just noise, the top fifty significant ICs are selected for the next calculation.

As shown in Figs 4 and 5, most of the signal energy is concentrated at the low frequency area, as the main differences lay between 2 and 20 Hz. The target event contains no salient difference from the base in the high frequency area; therefore, in the experiment, the DWT is used for eliminating the high frequency components, known as detail coefficients, of the EEG signal, and it lowers the data size to be managed. The low frequency components are left as the input for the next step of the process. Instead of a regular low-pass filter, using wavelet decomposition does not require the assumption that the spectrum is stationary throughout the whole segment, and it depicts spectrum changes along the time axis.

Since the details contain little difference between the target and base event, the DWT is applied first as a low-pass filter to remove the detail coefficients.

The energy of the wavelet coefficients is not used as a feature in this paper. The differences between the target and non-target event are not notable in the aspect of energy distribution (as seen in Figs 4 and 5), which cannot depict the difference of the signal itself. However, the waveforms exhibit clear differences; therefore, the WP coefficients implemented on selected ICs are used as a feature for classification, rather than the sub-band energy as usual.

For these reasons, the ICA is first applied to the EEG segments in one stimulation event. Then, the significant ICs are selected after the artificial ones have been removed. For each IC, the DWT is applied first as a low-pass filter to remove the detail coefficients, and then, a 3-layer WP is calculated on the approximate coefficients, and the first five sub-band coefficients generate one feature vector.

2.3. Classification

In order to determine to which group the images belong, there needs to be a classifier for the training classification model and for predicting the unknown events' labels based on the training-set attributes. Support vector machine (SVM) [10] has been widely used for high dimensional and non-linear classification.

SVM is a non-probabilistic binary classifier that is motivated by statistical learning theory and based on the structural risk minimization principle, and it has been regarded as one of the best machine learning methods. It has the ability to classify unseen patterns of the training set; therefore, it is widely used in the classification scheme. The SVM works out the hyper-plane that not only maximizes the separation margin between the two different classes but also minimizes the structure risk. In this way SVM can minimize the data fitting error and reduce the upper bound of the generalization error at the same time, thus increasing the generalization ability of the model.

The employment of the empirical risk minimization principle solves the over-learning problem that other machine learning algorithms might have. The training procedure is to solve a linearly-constrained convex quadratic programming problem, thereby avoiding the local minima problem. When the classification appears to be a non-linear one, a kernel function carries out mapping the samples into a higher dimension feature space, which avoids dimensionality disaster. There are many kinds of kernel functions, such as the Sigmoid function, the Polynomial, and the Radial Basis Function.

With the dataset $D = \{(x_i, y_i) \,|x_i \in R^p, y_i \in \{-1, 1\}\}_{i=1}^{k}$, the kennel function $K(x_i x)$, the Lagrange multipliers a^*, and the offset b^*, the objective function can be written as:

$$f(x) = sgn \left\{ \sum_{i=1}^{k} a_i^* y_i K(x_i x) + b^* \right\} \tag{2}$$

Besides the kennel function, the parameters also impact the classification result. The parameter C represents the penalty parameter of the errors. This parameter, which controls the trade-off between a large margin size and a small error penalty (which means large training errors), is chosen by the user. The larger C is, the larger the penalty is (and the smaller the training sample error) and the narrower the separation margin is. There is no systematic way of choosing C, and it is usually empirically selected by the user.

3. Experimental results and analysis

The MATLAB software suite was used for all data processions and computations; in particular, the toolkits FastICA [14] for ICA, ADJUST [13] for denosing, and Libsvm [15] for SVM were implemented.

As the relevant ERP's latency is usually between 300 ms and 500 ms from the stimulation onset, the EEG segments to be processed are cut up to 800 ms from the clip excitation onset. The dataset used in this paper includes 484 background events as well as 216 target events whose image clip contains an airplane target; 560 of which were chosen randomly for model training, and the rest were used for prediction. As the training data-set changes, the experiments were named as Dataset. 1$\sim$4 in Tables 1 and 2.

Table 1
Accuracy of classification under different features and datasets with default parameters

Feature	Accuracy			
	Dataset 1	Dataset 2	Dataset 3	Dataset 4
ICA mixing weights	19.6%	41.9%	71.4%	10.7%
WP sub-band energy	37.7%	59.3%	45.7%	56.2%
WP coefficients	59.6%	63.7%	61.6%	53.7%
ICA-WP	87.9%	90.0%	85.0%	87.9%

Table 2
Accuracy of classification under different parameters and datasets with ICA-WP

Weight (non-target: target)	Penalty factor	Accuracy			
		Dataset 1	Dataset 2	Dataset 3	Dataset 4
Default	Default	87.9%	90.0%	85.0%	87.9%
1:2	Default	87.9%	90.0%	85.0%	87.9%
1:2	10	88.6%	90.7%	85.7%	86.4%
1:20	10	88.6%	90.7%	85.7%	86.4%
1:2	100	87.9%	90.0%	85.0%	87.9%

3.1. Classification result

Table 1 gives the accuracies with different kinds of feature vectors as the input of SVM.

ICA separates signals into independent components; therefore, the mixing weights can represent the event signal's character. Under this kind of feature extraction, the ICA is performed on the training data, and the ICs and unmixing matrix are obtained. Then, the mixing weights, used as feature vectors, of predicting data are obtained by using the same unmixing matrix. Because of EEG's non-stationary character, the accuracy varies with different datasets.

The common usage of WP is the sub-band energy, which has a small-scale data amount and can notably reduce the training costs. As discussed in Section 2, the WP sub-band energy cannot indicate differences between the target and base stimulations, which emerges in the result.

The WP coefficients demonstrate a better performance as compared to the two above features. As the recordings of different channels are similar to each other, there is severe redundancy in the WP coefficients. Therefore, the efficiency of calculating features and the training model is very low. In addition, the large scale of the feature vector markedly harms the function of SVM.

Then, the ICA-WP performs best in these features, with the WP applied to chosen ICs and the feature generated from selected sub-band coefficients. This indicts that the ICA-WP can obviously reduce redundancy and adapt instability, which can extract the character and reduce the cost of calculation.

3.2. Parameter discussion

As shown in Table 2, the classification accuracy fluctuates between 85% and 90% as the training set is randomly chosen. This result indicates that ICA combined with WP decomposition, as a time-frequency analysis method, performs well in adapting the non-stationary and non-linearity of EEG and extracting the steady character.

In the training set, the amount of non-target events is about twice larger than that of target events, and as the SVM treats every sample equally, this results in more penalty quantum for a wrong prediction of non-target events than of target events. This condition affects the formation of separating the hyper-plane (i.e., trained classification model) and leads to the inclination to predict unknown events as non-target events. One simple adjusting method is to set a weight parameter to the types of training samples, thereby increasing the penalty of a wrong prediction of target events, which will counteract the effect of their small quantity. According to the results of the experiment, the weight adjusting does not change the accuracy even when the target weight is set to twentyfold. On one hand, this is because the imbalance in the training-set is mild and does not greatly affect the model. On the other hand, this indicates that the classifying ability of the extracted features is clarity, and the majority of samples are a certain distance away from the separating hyper-plane. The samples that might be wrongly classified are consistent, so the impact of the parameter adjustment on the accuracy is not observed.

The other factor that would affect the separating hyper-plane is the penalty parameter of the errors. The bigger the parameter is, the more effect the classifying model gets to distinguish the training set, as the wrong classification leads to more penalty. Meanwhile, with a model that has a less extended ability, the hyper-plane closes in on the training set. As the results show, with the same weight parameter and when the penalty factor is set to 10, the accuracy improves slightly; however, when the penalty factor is set to 100, most of the accuracies do not exhibit any more improvement. Since the large penalty parameter means there is a small separation margin and small training error, the result indicates that few samples are in the separation margin. It shows the classifying ability of the aforementioned feature extraction method in another way.

4. Conclusion

In this study, an algorithm that uses ICA-WP for feature extraction and SVM as a classifier is proposed to identify whether or not the image clips contain a target based on EEG signal processing. The results show that it is feasible to classify images by processing EEG signal segments obtained by a single trial, which is non-stationary and has low signal noise, rather than the average of multiple trials on the same image. ICA has certain advantages in multi-channel processing; it isolates independent ICs from brain activity and avoids redundancy among record channels. As a time-frequency method, WP can satisfy the non-stationary essence of brain activity so as to reflect the actual EEG signal spectrum. Combining with ICA and WP, the extracted features can express the signal well and are hardly influenced by changing the classifier's parameters.

With a feasible classification of EEG segments under a single image stimulation, it is possible to develop more applications to assist people with routine works, such as important area monitoring and warning and estimating satellite photographs, which requires people to identify targets from among a mass of images.

References

[1] G. Schalk, J. Mellinger, A Practical Guide to Brain-Computer Interfacing with BCI2000, London: Springer-Verlag, 2010.
[2] Peter Brunner, Anthony L. Ritaccio, et al. Rapid communication with a "P300" matrix speller using electrocorticographic signals (ECoG), Frontiers in Neuroscience, February 2011, Volume 5.
[3] Bogue R., Brain-computer interfaces: Controlled by thought, Industrial Robot, 37(2), 126-132, doi: 10.1108/0143991 10110188944.
[4] Squidoo, A brief history of mind control technology, 2010, http://www.squidoo.com/historyofmindcontrol.
[5] Hillyard SA, Imaging techniques: Event-related potentials (ERPs) and cognitive processing, New Encyclopedia of Neuroscience, 2009, 30: 13-18. doi: 10.1016/b978-008045046-9.00311-9.
[6] Linden DE, The P300: where in the brain is it produced and what does it tell us? Neuroscientist, 2005, 11: 563-576, doi: 10.1177/1073858405280524.

[7] Farwell, L.A., Donchin, E., Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials, Electroenceph Clin Neurophysiol, 1998, 70(6), 510-523.
[8] Yann, R., Lotte, F., OpenViBE: An Open-Source Software Platform to Design, Test, and use brain-computer interfaces in real and virtual environments, Presence, 2010, 19(1), 35-53.
[9] Hyvrinen A., The fixed-point algorithm and maximum likelihood estimation for independent component analysis, Neural Processing Letters, 1999, 10(1): 1-5.
[10] Vladimir N. Vapnik, The nature of statistical learning theory second edition[M], New York, Spring-Verlag.
[11] Mallat S., A Wavelet Tour of Signal Processing, Second Edition, Beijing, China Machine Press.
[12] Nima B.S., et al., EEG dataset from Swartz Center, UCSD, http://headit-beta.ucsd.edu/studies.
[13] Andrea Mognon, et al., ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features, Psychophysiology, 48 (2011), 229-240.
[14] Hugo G, et al., FastICA package, an ICA program implementing fast fixed-point algorithm, http://www.cis.hut.fi/projects /ica/fastica/.
[15] Chih-Wei Hsu, et al., LIBSVM package, a library for Support Vector Machine, http://www.csie.ntu.edu.tw/∼cjlin/ libsvm.