Brain Topogr (2015) 28:33–46
DOI 10.1007/s10548-014-0371-9

ORIGINAL PAPER

Decoding Objects of Basic Categories from Electroencephalographic Signals Using Wavelet Transform and Support Vector Machines

Mitra Taghizadeh-Sarabi · Mohammad Reza Daliri · Kavous Salehzadeh Niksirat

Received: 6 October 2013 / Accepted: 27 April 2014 / Published online: 17 May 2014

Abstract Decoding and classification of objects through task-oriented electroencephalographic (EEG) signals are the most crucial goals of recent researches conducted mainly for brain–computer interface applications. In this study we aimed to classify single-trial 12 categories of recorded EEG signals. Ten subjects participated in this study. The task was to select target images among 12 basic object categories including animals, flowers, fruits, transportation devices, body organs, clothing, food, stationery, buildings, electronic devices, dolls and jewelry. In order to decode object categories, we have considered several units namely artifact removing, feature extraction, feature selection, and classification. Data were divided into training, validation, and test sets following the artifact removal process. Features were extracted using three different wavelets namely Daubechies4, Haar, and Symlet2. Features were selected among training data and were reduced afterward via scalar feature selection using three criteria including T test, entropy, and Bhattacharyya distance. Selected features were classified by the one-against-one support vector machine (SVM) multi-class classifier. The parameters of SVM were optimized based on training and validation sets. The classification performance (measured by means of accuracy) was obtained approximately 80 % for animal and stationery categories. Moreover, Symlet2 and T test were selected as better wavelet and selection criteria, respectively.

Keywords BCI · EEG signals · Object recognition · SVM classification · Wavelet

M. Taghizadeh-Sarabi
Virtual Center, Iran University of Science and Technology (IUST), Tehran, Iran

M. R. Daliri (✉)
Biomedical Engineering Department, Faculty of Electrical Engineering, Iran University of Science and Technology (IUST), 16846-13114 Tehran, Iran
e-mail: daliri@iust.ac.ir

M. R. Daliri
School of Cognitive Sciences (SCS), Institute for Research in Fundamental Sciences (IPM), Niavaran, 19395-5746, Tehran, Iran

K. S. Niksirat
Department of Control Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran

Introduction

The human brain is a complex system in which neural activity generates rich encoded information. The neural information can be used to investigate encoding in the brain and decoding the stimuli from brain signals. One of the important cognitive tasks in the brain is object recognition. The procedure of visual object perception is yet to be well-discovered; however, a number of studies have made substantial progress (Johnson and Olshausen 2003; Martinovic et al. 2011). To allow studying object recognition, several non-invasive recording techniques of neural responses have been allocated. Among the noninvasive techniques, brain activity can be inferred from electroencephalograms (EEG). The EEG is a well-documented technique which is capable of characterizing certain brain states, especially processing of different semantic categories (Hoenig et al. 2008; Pulvermuller et al. 1999; Kiefer 2001; Paz-Caballero et al. 2006; Proverbio et al. 2007; Fuggetta et al. 2009; Adorni and Proverbio 2009).

EEG signals are mainly analyzed by their frequency contents. That is, the interpretation of the EEG signal is based on the power of its constituting frequencies. Five main frequency ranges are normally included in all EEG studies: delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (from 30 Hz). The development of EEG-based interpreting approaches is an interesting application which makes real-time decoding systems possible (Muller et al. 2008). In order to decode the mental tasks, three main aspects can be considered to analyze EEG signals namely feature extraction (Sykacek et al. 2003; Ince et al. 2005), feature selection (Pregenzer and Pfurtscheller 1999; Lal et al. 2004), and classification approaches (Palaniappan et al. 2002; Peters et al. 2001).

The feature extraction stage involves transformation of raw signals into relevant data structures by deleting noise and highlighting important data. The products of this process are called feature vectors. In addition, it could imply "dimensionality reduction", which eliminates redundant data from the feature vectors with the aim of facilitating the classification process. Moreover, extracted features should be unique to each data class and be different of the other class (Coyle et al. 2005, Coyle et al. 2006a, b; Wolpaw et al. 2002). One of the approaches of feature extraction is fourier transform (FT). The Fourier transform converts a signal in the time domain to a signal in the frequency domain, and it is widely used in the form of the fast fourier transform (FFT) algorithm. Fourier analysis is simply not effective when used upon non-stationary signals, because it does not provide frequency content information localized in time. The short time fourier transform (STFT) was developed for analyzing a small section of the signal at a time called "windowing". The problem with STFT is the resolution. With a smaller window size, the faster changing components are better detected, while slowly changing details are not picked up well. Lower frequencies can be detected with larger windows, but localization in time will not be suitable in this case (Messer et al. 2001). To solve the drawbacks of power spectrum methods such as FFT and STFT, wavelet transform (WT) has been proposed to perform time-scale analysis of signals. The major advantage of WT over spectral analysis methods is that it is appropriate for analysis of non-stationary signals; therefore, it is proper to localize transient events. According to the non-stationary nature of biological signals, wavelet's feature extraction and representation methods can be used to analyze them. Wavelets are powerful candidates for decomposition and feature extraction of EEG signals for many applications due to their multi resolution temporal and spectral locality (Sherwood and Derakhshani 2009).

Messer et al. (2001) studied a set of wavelet families and level of decompositions that acted best in removing noise in phonocardiogram (PCG). They used discrete wavelet transform (DWT) to calculate the wavelet's coefficient at discrete intervals of time and scale instead of continuous wavelet transform (CWT), which reveals more details about the signal, but with much more computation time. Many factors that must be considered when attempting to de-noise psychological signals when using wavelets. One of them is orthogonality. Being computationally inexpensive, orthogonal wavelets allow a transform to be computed containing the same number of points as the original signal. Wavelets having properties of orthogonality include Haar, Daubechies, Coiflets, and Symlets. An increasing level of wavelet requires increasing complexity and more computation time. In most cases, using 5 levels of decompositions proved adequate.

Due to properties of bio-signal types, allocating proper mother wavelet is essential for obtaining better performance. Rafiee et al. (2011) examined more than 300 wavelets on three categories of biological signals [electromyogram (EMG), vaginal pulse amplitude (VPA), and EEG signals]. They recorded EEG from three mono-polar electrodes during visual stimulus presentation. They found that daubechies44 (Db44) is the most similar function to biomedical signals. Unlike previous researches, Rafiee proved that asymmetric wavelets such as Db44 can have better results than symmetric wavelets. They also proved that similarity between signals is not always suitable for signal processing based on wavelet transform. In another study Daubechies6, Daubechies1, Daubechies2, Symlet6, Symlet10, Coiflets4, and Coiflets2 wavelets with learning vector quantization–wavelet transformer (LVQ–WT) and multilayer perceptron–wavelet packet (MLP–WT) classification methods were used by Keshtiban et al. (2011). Best results were obtained from Daubechies and Symlets wavelets.

Krishnan Mookiah et al. (2012) in their study extracted features for identification of normal and glaucoma classes using higher order spectra (HOS) and DWT. Then, extracted features were fed to the SVM classifier with linear, polynomial order 1, 2, 3 and radial basis function (RBF). SVM classifier with kernel function of polynomial order 2 achieved an accuracy of 95 % sensitivity and specificity of 93.33 and 96.67 % respectively to identify the glaucoma and normal images.

Gu et al. (2009) performed offline single-trial EEG classification to identify the speed of an imagined wrist extensions movement at two speeds (fast and slow) from EEG recordings in amyotrophic lateral sclerosis (ALS) patients. They implemented DWT and SVM, and proved that speed of the task was encoded in the time delay of peak negativity, which was shorter for faster movements. In another similar study, Demiralp et al. (1999) employed a response-based classification procedure to choose the trials containing the P300 component from the whole set of sweeps of an auditory oddball paradigm. For this purpose, the most significant response property reflecting the P300 wave was identified by using the WT. The average of selected single sweeps from the whole set of data according to this criterion yielded more enhanced P300 waves compared with the average of the target responses.

Category recognition is a challenging problem and techniques based on computer vision often require human involvement to learn good object category models. Since the computer should learn how to recognize a given EEG pattern, correspondence among EEG patterns and computer actions involves a machine-learning problem. A computer algorithm is in charge of extracting the associated EEG patterns. After the training phase is finished, the subject should be able to control the computer actions with his/her thoughts. This is the major goal for a BCI system. Corresponding to this aim, (Philiastides and Sajda 2005) demonstrated neural measurements of perceptual decision making via single-trial EEG analysis in a face versus car categorization task. They identified two major separating components. These findings demonstrate a temporal distribution of component activity showing an evidence gathering process, which begins after early visual perception. Misaki et al. (2010) attempted to decode the category of visual objects information in 3-Tesla functional magnetic resonance imaging (3T-fMRI) response patterns measured for approximately 60 min per subject with six multivariate classifiers. They proved Fisher's linear discriminant and the linear support vector machine performed best and suggested that linear decoders may execute most successfully in the mentioned scenario of visual object representations. In comparison, in another study, Garrett et al. (2003) compared the result of a linear classifier (Linear Discriminant Analysis) and two non-linear classifiers (NN and SVM) applied to the classification of spontaneous EEG during five mental tasks. The results indicated that nonlinear classifiers produce better classification performance, in which the performance of non-linear SVM was 72 %, while linear discriminant analysis performance was near 66 %.

Tzovara et al. (2012) exhibited neuro-imaging experimental conditions using single-trial EEG responses to decode stimulus-related signals in two event related potential (ERP) studies. They used statistical distribution method with a gaussian mixture model (GMM). Cross-validation algorithm was tested in two independent EEG datasets to classify left versus right hemi-field checkerboard stimuli for upper and lower visual hemi-fields, and in an initial versus repeated presentations of visual objects. Martinovic et al. (2008) in their object recognition study proved that object's features coded very rapidly and play different functional roles while color, extra contours and edges delay it.

In the current study, features were extracted using DWT with three mother wavelets, Haar, Symlet2, and Daubechies4 to perform the EEG single-trial pattern classification. Scalar feature selection approaches were applied to reduce the feature set dimensionality through selecting a subset of features. Feature ranking with three different criteria (*T* test, Entropy, and Bhattacharyya distance) were performed (Chang et al. 2010). It is essential to reduce computational cost and improve classification performance, especially when dealing with finite sample sizes. Finally, the selected features related to EEG patterns were classified using SVM with RBF kernel, sequential minimal optimization (SMO) solver and one-against-one model.

Methods

All subjects were informed of the task prior to experiment. They were gifted for their participation and signed written consent forms in accord with the declaration of Helsinki principles. The study was approved by the National Committee of Ethics in Medical Research (Technology and Research Deputy of Ministry of Health and Medical Education).

Participants

A total of ten adult volunteers (8 male and 2 female, age range 18–28; mean age: 23 $\pm$ 3.4 SD), participated in the study. All participants were right-handed except one, and reported that they did not suffer from any psychological or neurological disorder, and had normal vision.

Stimuli

Same-sized (600 $\times$ 800 pixels) color images of 12 categories were taken from the internet and included images of animals, flowers, fruit, transportation devices, body organs, clothing, food, stationery, buildings, electronic devices, dolls, and jewelry. The experiment was displayed in two parts. Six categories were presented in partition-1 and the next six categories in partition-2, which can help to reduce participant's eye boredom and EEG eye-blinking artifacts. Each category had five different items. Each subject observed 360 images during the whole experiment. Subjects were allowed to have 3 min rest between two parts. Categories were not randomly assigned to partition-1 and partition-2 and have had fixed order for all subjects. Moreover, each trial (given category) consisted of 30 images, 15 of which were related to the given category and the remaining 15 images were from others, which were non-targets. Non-target images were randomly selected among other categories and were shown within the stream of targets. The trials were presented only once per participant with no information about their order. Figure 1 shows task paradigm, and orders of pictures are shown in Fig. 2.

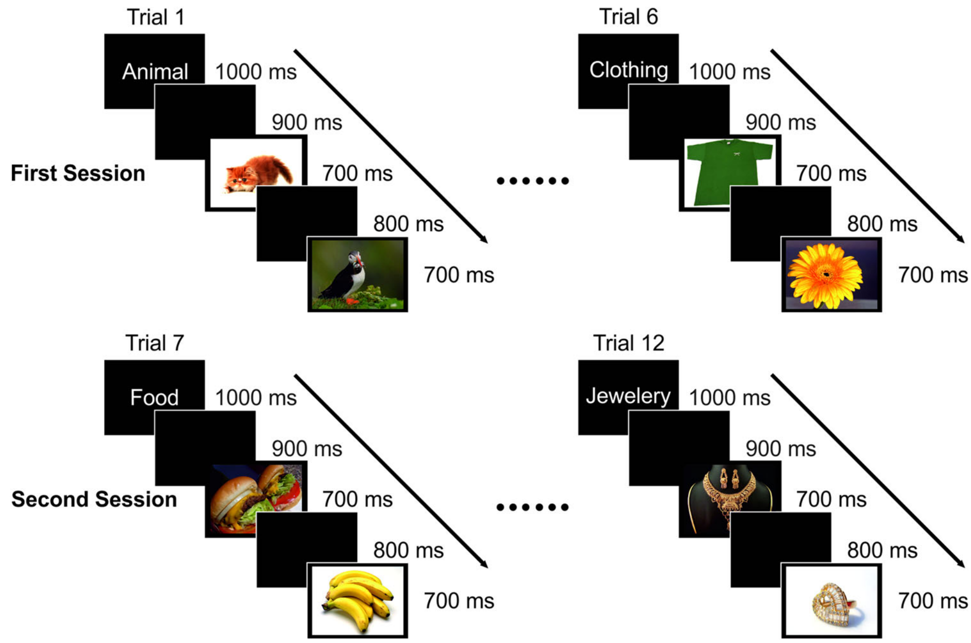

Fig. 1 "Cued-target" with "go/no-go" task paradigm. Target cue remained on the screen for 1 s, after an additional 900 ms delay, the images were presented. Each image was presented for 700 with 800 ms interval time between two images and 5 s *black screen* for rest between two different categories

Experimental Design

The defined experiment in this study is "cued-target" with "go/no-go" task. In the "go/no-go" task, subjects press the button if the image contains a target and do nothing if the image is not a target. In the cued-target, a word (target cue) was presented on the screen before each category, informing the subjects to target category. Target cue remained on the screen for duration of 1 s, and after an additional 900 ms delay, the images are presented. Each image is presented 700 ms throughout 800 ms interval time between two images. A black screen is displayed for 5 s between two categories as a rest time. Subjects were asked to click the left button of the mouse with their right hand index finger to make responses as quickly as possible. Moreover, they were asked to delay their blinks in 800 ms interval time. All images were centrally presented on a LCD monitor with a viewing distance of 75 cm. Image presentations were controlled by a PC running Psytask–WinEEG presentation software. PSYTASK is software for audio/visual stimuli presentation, which works with WinEEG, providing a synchronous stimuli presentation to EEG recording. The particular time delays between image presentations were chosen by PSYTASK.

EEG Recording

Participants were fitted with a 19-channel electrode cap and prepared for EEG recording according to standard techniques. Recorded channels (FP1, FP2, F3, F4, C3, C4, P3, P4, F7, F8, T3, T4, T5, T6, FZ, CZ, PZ, O1, and O2) were selected among the international 10–20 set of electrode positions with linked-ears montage (Miller et al. 1991). (The MCN system (modified combinatorial nomenclature) renames four points of the 10–20 system T3, T4, T5, and T6 as T7, T8, P7, and P8, respectively.) Subjects performed the experiment in a darkened, sound-dampened, electrically shielded booth. EEG signals were amplified with MITSAR hardware, and then sent through an analog-to-digital converter. Signals were recorded at 500 Hz sampling frequency on a PC running Digitize.

Preprocessing

Brain signals should be preprocessed to achieve better analysis performance. Due to the most useful frequencies of brain bands, brain signals were filtered between 1 and 30 Hz frequencies (Simanova et al. 2010). A 1–30 Hz phase-shift free butterworth band-pass filter (12 dB/Octave) was used. Moreover two amplitude thresholds were designed for fast and slow waves in which voltages higher than 50 µV for slow waves and 30 µV for fast waves were rejected using MATLAB/signal-processing-toolbox [threshold values were chosen based on Alpha and Beta brain waves normal amplitudes (Sanei et al. 2007)].

In order to correct detailed artifacts, the ICA method of WinEEG was implemented. The "Infomax" algorithm was implemented in WinEEG software to analyze raw EEG signals (Delorme et al. 2007). Eye blink artifacts and some others can be corrected using ICA method even if the EOG signal was not recorded. This method is based on blind source separation procedure between multi-channel EEG data and spatial filtering of some components of the EEG signal. Input data is manually selected during time interval including

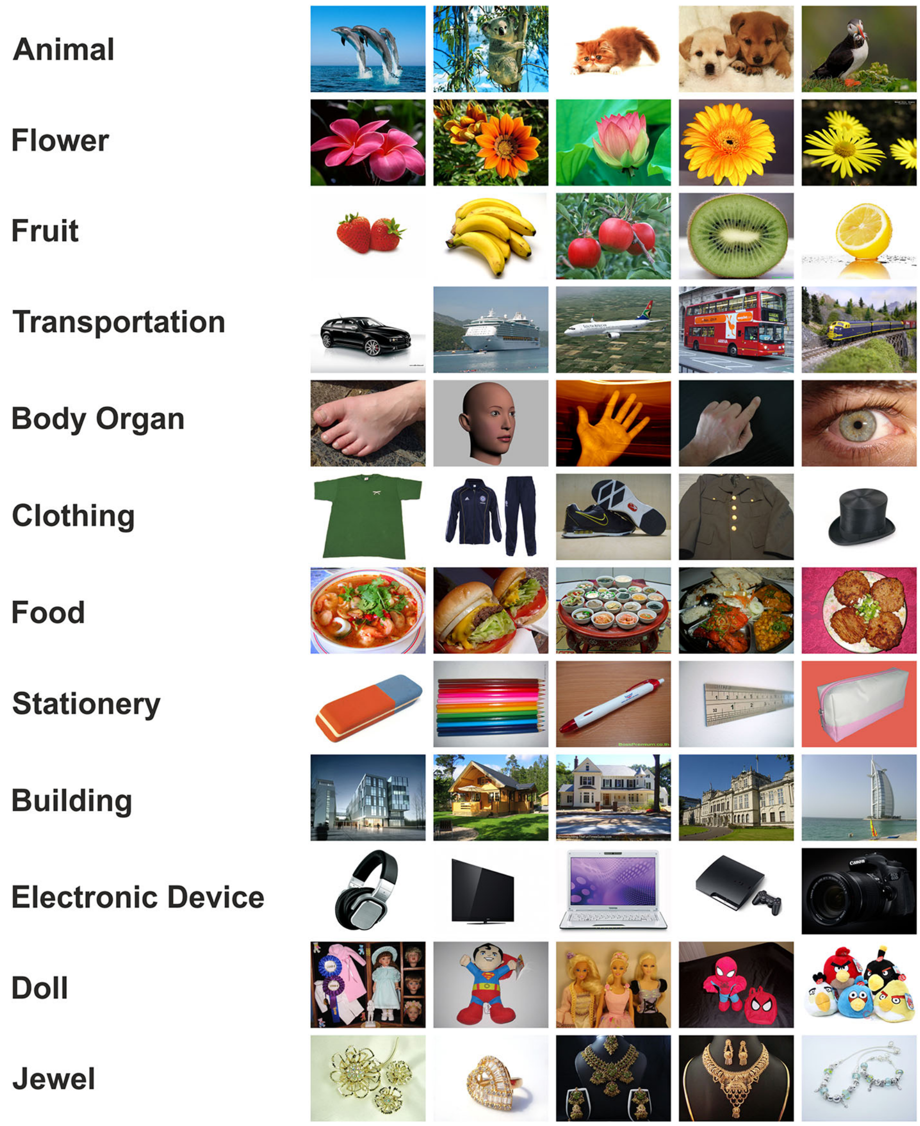

Fig. 2 The images of study. Twelve categories 600 × 800 pixels color images including images of animal, flower, fruit, transportation device, body organ, clothing, food, stationery, building, electronic device, doll and jewel

artifacts. After decomposition of the multi-channel signal, components of the signal related to artifacts were selected manually during the analysis of their topographies and waveforms of components. In general, the main components were horizontal and vertical eye movements and temporal muscular activities, so these noisy components were selected and ICA algorithm was applied to the whole EEG data.

The components corresponding to artifacts were removed and a spatial filter performing equivalent transformation was calculated and applied to raw EEG (Jung et al. 2000). Each trial (given category) had fifteen target images; however, this number is not equal for all categories and all subjects for the reason of removing and cancelling wrong and noisy responses. In order to have the same number of images for all categories, thirteen target responses were chosen, which results equal training probability. Continuous EEG signals were divided into segments. The time interval of 40 ms from the picture onset was removed, and the remaining 660 ms were used for analyzing each image. The 0–40 ms time bands were rejected from the analysis, because it takes approximately 40–50 ms for information to reach the primary visual cortex (Phillips et al. 2012). This 660 ms signal converted to 330 samples due to 500 Hz sampling rate. Considering the 19 EEG channels, the dimension of one single trial equals 19 × 330.

Data from 8 subjects were preprocessed and used for classification because the 4th and 5th participant data were full of uncorrectable artifact. Therefore, they were rejected for further analysis.

EEG Feature Extraction

Wavelet transform is a spectral assessment procedure in which any function can be declared as an infinite series of base wavelets. Wavelets are mathematical functions that decompose data into different frequency components by shifting and scaling, which lead to a set of signal coefficients. The signal can be reconstructed by these coefficients. The main feature of WT is time–frequency localization. In EEG signals, the wavelet method will exhibit features related to the temporary character of the signal, which obviously cannot be seen by the fourier transform (Subasi 2005; Kandaswamy et al. 2004). Wavelet transform can be expressed in terms of a low-pass filter and a high-pass filter.

Wavelet analysis comprises CWT and DWT. DWT is a more beneficial procedure than CWT. DWT analyzes the signal at different frequency bands decomposing the signal into an approximation and detail data. Each step of the scheme involves two digital filters with two down-sampling blocks by 2. Outputs of the first high-pass and low-pass filters generate two decompositions named details (D1) and approximation (A1), respectively. Additionally, this process will continue hierarchically and A1 is extra decomposed and continued.

Applying appropriate mother wavelet and the levels of decomposition are essential factors in signals analysis. The level of decomposition is selected based on the dominant frequency components of the signal. In the present study, the number of decomposition levels was chosen to be 5, because the EEG signals were filtered between 1 and 30 Hz (Simanova et al. 2010; Sanei et al. 2007). Thus, the EEG signals were decomposed into details D1 to D5 and one final approximation, A5. These are type of energy coefficients, which are derived from each electrode. Tests were performed with different types of wavelets (Symlet2, Haar, and Daubechies4) (Merry 2005). N is the wavelet order. Wavelet orders for Haar, Sym2, and Db4 are 1, 2, and 4, respectively. The length of each filter is equal to 2 N. The length of Approximation and details is equal to:

$$\begin{cases} \text{Length }(D_i) = \left\lfloor \dfrac{\text{Length }(D_{i-1}) + 2N - 1}{2} \right\rfloor_{i=1,\ldots,5} \\ \text{Length }(A_i) = \left\lfloor \dfrac{\text{Length }(A_{i-1}) + 2N - 1}{2} \right\rfloor_{i=1,\ldots,5} \\ \text{Length }(A_0) = \text{Length }(D_0) = 330 \end{cases} \tag{1}$$

where i is the decomposition level. For instance in Db4, the length of D1 is equal to $\left\lfloor \frac{330+(2\times4)-1}{2} \right\rfloor = \lfloor 168.5 \rfloor = 168$. Sequentially, lengths of D2, D3, D4, D5, and A5 are 87, 47, 27, 17 and 17, respectively. Table 1 shows length of other wavelet's decomposition.

Feature Selection

In order to reduce the number of features easily, all decompositions were normalized and reshaped into a row vector. Table 1 shows the applied wavelets with the number of extracted reshaped features. Notice that the number of features varies based on wavelet type. For example, in Haar wavelet, which has a basic shape, the number of features is less than the Daubechies wavelet. In order to select essential and proper features for classification, several scalar feature selection methods were implemented using different criteria including T test, entropy, and Bhattacharyya distance, which they ranked all features. T test criterion returns the significance level (p value) of the test. The p value is the probability, under the null hypothesis, of observing a value as extreme as or more extreme than the test statistic. The Bhattacharyya distance has been used as a class separability measure for feature selection, and is known to provide the upper and lower bounds of the Bayes error. However, the bounds are not tight enough for practical applications (Choi and Lee 2003).

According to early discussed methods, the top 1,000 features were selected according to each criterion. The aim of this initial selection is to reduce the complexity at the first step. The absolute value of the criterion was used to rank features. Absolute value is related to the criterion used in feature selection, which means how much a feature is significant to separate two classes. Features with high absolute value were chosen and others were rejected. The scalar feature selection method considers feature ranking between the two classes and due to twelve categories in the current study, the mentioned method was applied 66 times considering all states ($\binom{12}{2} = \frac{12!}{10!\times2!} = 66$). The matrix size 66 × 1,000 was obtained. Finally, the features with most repetitions were selected. The most repeated features means the most informative features, which have a high count number among 66 states. In other words, the most informative features are those which exist in most 66 states. All 66 × 1,000 ranked features were counted, and features with more existence in most states were chosen as the best features.

Classification

Support Vector Machines

A support vector machine (SVM) is generally a binary classifier based on minimization of structural risk. In complicated nonlinear modeling, the SVM maps the training data into a higher dimensional feature space in which a linear hyper-plane can separate the data (Vapnik 1998). In the simplest form (binary linearly separable case), the SVM training phase tries to find the linear function to separate the data:

Table 1 Number of decomposed features of one image using DWT with three different wavelets

Wavelet	Haar	Symlet2	Daubechies4
D1	19 × 165	19 × 166	19 × 168
D2	19 × 83	19 × 84	19 × 87
D3	19 × 42	19 × 43	19 × 47
D4	19 × 21	19 × 23	19 × 27
D5	19 × 11	19 × 13	19 × 17
A5	19 × 11	19 × 13	19 × 17
Total features	19 × 333 = 6,327	19 × 342 = 6,498	19 × 363 = 6,897

19 × 330 samples per image decomposed into 19 × 333, 19 × 342 and 19 × 363 samples by Haar, Symlet2 and Daubechies4 wavelets, respectively

$$f(x) = \omega^T x + b \tag{2}$$

where ω is the weight vector, x is the input data, and b is considered as bias. The SVM classifier is capable of finding possible hyper-planes among several numbers of hyper-planes, which maximizes the margin between two classes using support vectors at closest point to the hyper-plane. Minimizing the following cost function leads to the optimal hyper-plane:

$$J(\omega) = \frac{1}{2}\omega^T\omega = \frac{1}{2}\|\omega\|^2 \tag{3}$$

In the non-linear case as considered above, the idea is to map the input data to some higher dimensional space, where the data can be classified linearly. The following mapping can be used:

$$\Phi: \quad R^N \longrightarrow R^M \tag{4}$$

where N is the dimension of input space, and M is a higher dimension space, termed feature space. In feature space, the technique described above can be used to find an optimal separation hyper-plane.

Multi-class SVM

One of the extended SVM for solving multiclass problems is the "one-against-one" strategy, which constructs one SVM for each pair of classes. Thus, for a problem with c classes, $c(c-1)/2$ SVMs are trained to distinguish the samples of one class from the samples of another. The Max Wins algorithm performs voting among the classifiers and the selected label is the one with the most votes among the classifiers (Kressel 1999). In case that two classes have identical votes, though it may not be a good strategy, simply the class appearing first in the array of storing class names was chosen (Chih-Wei and Chih-Jen 2002). In addition, in cases with more classes, the "one-against-one" strategy is more accurate than other methods, and the unbalance of the number of the samples, especially when it has few training samples per class does not cause any problem (Milgram et al. 2006).

In the current study, the "one-against-one" multiclass SVM was used to classify 12 object categories using radial basis function (RBF) kernel and sequential minimal optimization (SMO) solver. The proper kernel function choices can remarkably enhance the system's performance and decrease computational cost. Besides the linear kernel, which is the simplest form of kernel, the prevalently used kernel is Gaussian kernel, which is defined in (4). In this kernel, σ controls the width of the Gaussian kernel. The RBF kernel is a reasonable choice. This kernel nonlinearly maps samples into a higher dimensional space, so unlike the linear kernels, it can handle the case where the relation between class labels and attributes is nonlinear (Hsu et al. 2010).

$$K(x_i x_j) = e^{-\|x_i - x_j\|^2 / 2\sigma^2} \tag{5}$$

In consequence to solve the optimization problem for training SVM an effective algorithm should be used. The algorithm used for our experiments is called SMO, which is an iterative approach. In this approach, a series of smallest possible sub-problems are created from the problem, and an analytical solution is provided for these sub-problems (Bottou et al. 2007).

Significance Test

For a statistically significant test, random sub-sampling or Monte Carlo cross validation were used. This method randomly splits the dataset into training and test data without overlapping. For each such split, the model is fitted into the training data, and predictive accuracy is assessed using test data. The results are then averaged over the splits. Each time, training and testing data were selected randomly to achieve different performances. The mentioned process was repeated 10 times, and the final performance was averaged and reported. Raw data was randomly divided into two training and testing data, 70 % for training, and 20 % for test data. Moreover 10 % of the data was left for validation; this part will be used for setting SVM parameters (Pereira et al. 2009).

Results and Discussion

Eight volunteers' EEG signals of twelve image categories were classified using SVM individually during visual stimuli task and an overall 70 % classification performance was obtained. In previous studies, researchers had attempted to classify categories in different BCI applications such as identifying objects from their spoken name, visual representation, and written name in ERP study with 89 % performance (Simanova et al. 2010), left or right imaginary hand movement with 82–92 % classifier accuracy (Higashi et al. 2009; Zhang et al. 2008), and so on. The most previous related papers are designed and defined in maximum three or four classes, however our work is much more complicated consisting of 12 classes, and naturally, it will have less performance accuracy than previous studies, because the confusion probability is higher. Moreover, single trial classification has not been well addressed up to now. In this paper, all tests were carried out using MATLAB and WinEEG software.

SVM Parameters

The effectiveness of SVM depends on the selection of kernel, the kernel's parameters σ, and soft margin parameter C (regularization constant). A common choice is a Gaussian kernel (RBF), which has a single parameter (Kernel Argument). The regularization constant parameter controls the trade-off between errors of the SVM on training data and margin maximization. The feature selection method discussed earlier was implemented only on training data. As we mentioned before, data were divided into training, validation, and testing parts. SVM was designed and trained based on training features, and was then tested. Regularization constant and kernel argument parameters of SVM were set up depending on the training data. However, these parameters were optimized using validation data. The SVM parameters were setup by "leave one out cross validation" without any interference of test data by only using training and validation data. This method was repeated 10 times. The goal is to identify optimal regularization constant and kernel arguments. A grid-search on regularization constant and kernel arguments using leave one out cross-validation was carried out. Various pairs of regularization constant and kernel argument values were tried, and the one with the best cross-validation accuracy was picked up. Optimum SVM parameters were calculated and set separately for each subject. Figure 3 shows grid-search over SVM parameters only for subject 1, which [1000, 1] is the best choice with 0.19 validation errors for kernel argument and regularization constants, respectively.

Image Presentation

In our experiments, we displayed the name of the target categories before presenting pictures to inform participants about images that should be classified by left clicking the mouse. In comparison with (Johnson and Olshausen 2003), the task of the study was cued-target and single-category, and compared with each other using ERP, although our task is similar to cued-target with some difference in presentation timing. On the other hand it has other standpoints of classification with mathematical and machine learning algorithms. All stimuli were selected colorful pictures, which are used in real life objects. It has been shown that colorful pictures lead to faster responses rather than line drawing pictures (Martinovic et al. 2011). In order to diminish eye movements, tiredness and consequently eye related artifacts, 12 categories were divided into 2 six-groups within a 3 min time gap between (Martinovic et al. 2011).

Feature Selection and Extraction Effects

The effects of feature selection and feature extraction methods should be evaluated before final evaluation. Three categories were randomly examined (animal, stationery, and jewelry). Preprocessed raw signals were classified without using wavelet transform. Whole results were obtained with very low performance about 32 %. However, after applying Sym2 wavelet to the same three categories, 27 % increase in performance was observed and reached 59 % ($F(1,7) = 31.6$, p value < 0.005). In the next step, feature selection method with T test criterion applied and dimensionality reduced features were classified with 87.3 % performance ($F(1,7) = 47.4$, p value < 0.005). In fact, increasing the number of categories leads to performance decrease. Notice that F and p represent statistic and significance level, respectively by one-way repeated measure ANOVA test.

Optimal Number of Features

The purpose of feature selection is to reduce feature dimensionality. Classification performance is directly related to the number of features and criteria selection. Figure 4 illustrates average of 10 classification performances (tenfold Monte Carlo cross validation) of subject 1. Results indicate an optimum number of features to T test, entropy and Bhattacharyya distances with 260, 400 and 600, respectively. Each subject had different optimum feature numbers with each other. Results demonstrate that the T test criterion needs fewer numbers of features for the same performance among other criteria. This is also true in the case of other seven subjects. Additionally, low number of features led to fast training. There is a significant difference between numbers of features; therefore, T test was chosen for analysis. The selected features were classified using support vector machine with one against one model,

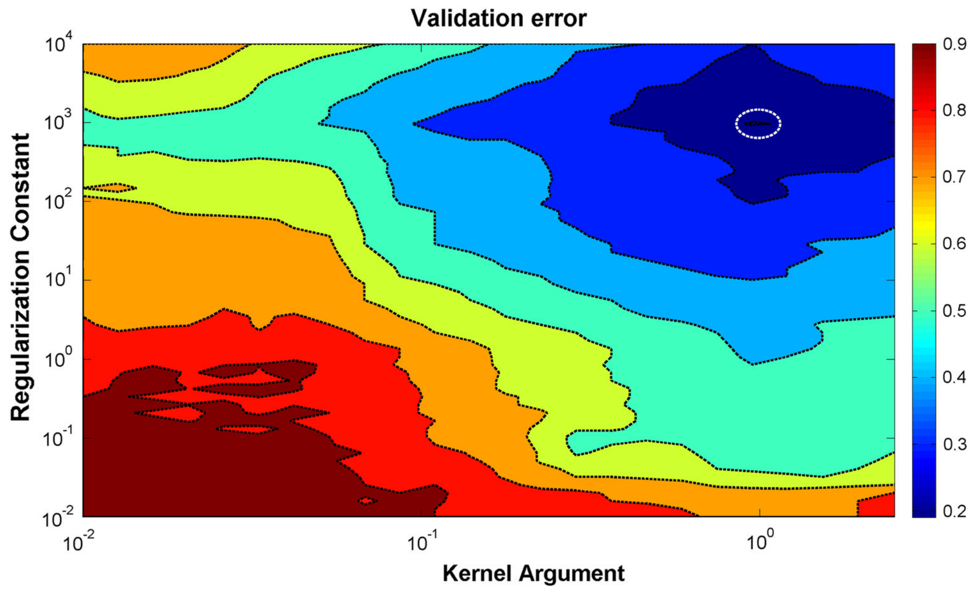

Fig. 3 Kernel argument and regularization constant. A grid-search on regularization constant (C) and kernel argument (σ) using leave one out cross-validation was done for subject 1. Various pairs of regularization constant and kernel argument values are tried and the one with the best cross-validation accuracy is picked. This figure shows (1000,1) is best choice with 0.19 validation error for kernel argument and regularization constant, respectively. Optimum SVM parameters were calculated and set separately for each subject

RBF kernel, and SMO solver. All data were decomposed into training, validation, and testing sections; 70 % for training, 10 % for validation, and 20 % for testing. Results were repeated 10 times, with random training and test dataset selection.

Best Combination of Wavelets and Criteria

In order to evaluate criteria and wavelets, a performance index for each combination was inspected at Table 2. It can be seen that T test criterion and Haar wavelet present the best results among others; however, Sym2 with T test is the best combination. The averaged performances of criteria were entered in the last column of the table, and the last row of the table indicates the averaged performances of wavelets. Additionally, different feature extraction methods and different feature selection criteria were compared to check the statistical significance in the performances. For the statistical test, a two-way repeated measure ANOVA was used. The results indicated that there was the significant effects of the wavelet type ($F(2,14) = 509$, p value < 0.005), the feature selection type ($F(2,14) = 218.5$, p value < 0.005) and the significant interaction between them ($F(4,28) = 32.8$, p value < 0.01). We performed the paired-samples T test between each pair of conditions too. It should be noted that the results have very similar mean values, but the standard errors of the mean are very small, and there is no overlap between the results. For the effect of the wavelet types, by fixing the T test as the feature selection method, which has the best results for this part, the results of the paired-samples T test show that there was a significant difference in the scores for the Sym2 wavelet ($M = 70.03$ %, SE $= \pm 0.044$) and the Db4 ($M = 67.38$ %, SE $= \pm 0.046$) conditions ($t(7) = 45.43$, p value < 0.01), and also significant differences in the scores for the Sym2 and the Haar ($M = 69.19$ %, SE $= \pm 0.030$) conditions ($t(7) = 19.53$, p value < 0.01). These results suggest that the Sym2 wavelet performs significantly better than the Db4 and the Haar. For the effect of feature selection types, we applied the statistical test over the performances of the feature selection methods in pair by fixing the Sym2 wavelet type, which provides the best results for the feature extraction part. The results of the paired-samples T test show that there was a significant difference in the scores for the T test criterion ($M = 70.03$ %, SE $= \pm 0.044$) and the entropy ($M = 66.78$ %, SE $= \pm 0.047$) conditions ($t(7) = 67.05$, p value < 0.01), and also significant differences in the scores for the T test and the Bhattacharyya ($M = 64.46$ %, SE $= \pm 0.054$) conditions ($t(7) = 80.09$, p value < 0.01). These results suggest that the T test criterion works significantly better than the Entropy and the Bhattacharyya. Notice that t, M and SE represent t value, mean and standard error, respectively.

Three one-way repeated measure ANOVAs were implemented to study whether there was a significant effect of wavelet type on performances for each feature selection type independently. The results reveal the significant effect of the Sym2 ($F(2,14) = 57.8$, p value < 0.01), the Db4 ($F(2,14) = 32.5$, p value < 0.01), and the Haar ($F(2,14) = 7.5$, p value < 0.01) on performance changes depending on the level of feature selection criterion. On the other hand, three one-way repeated measure ANOVAs were applied to explore whether there was a significant effect of feature selection criterion on performances for each wavelet type independently. The results indicate the significant effect of the T test ($F(2,14) = 18.3$, p value < 0.01), the Entropy ($F(2,14) = 6.4$, p value < 0.01), and the Bhattacharyya ($F(2,14) = 17.2$,

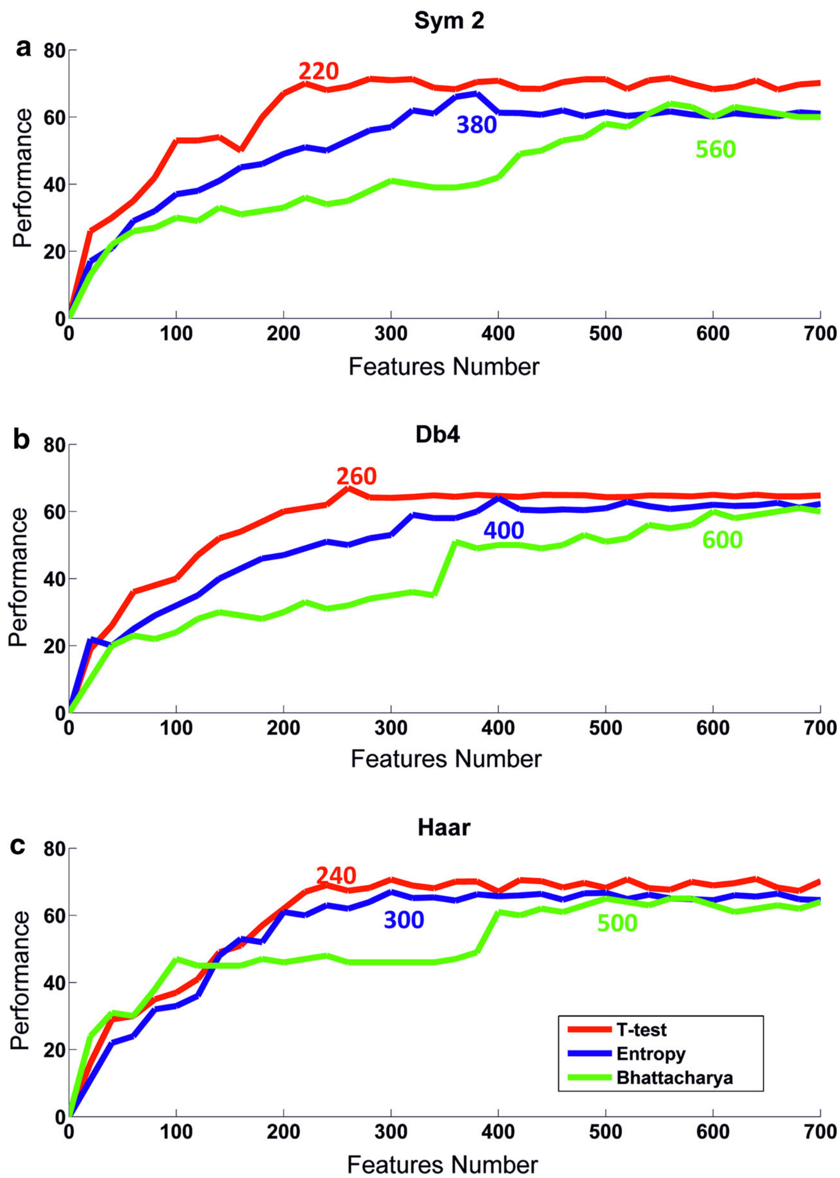

Fig. 4 Performance comparison for different number of features consists of three wavelets and three criteria. X and Y axis indicate feature number and performance, respectively. Results are average of ten classification performances (tenfold Monte Carlo cross validation) of subject 1. Each subject had different optimum feature numbers with each other. **a** The number of features for Sym2, (*T* test, entropy and Bhattacharyya distance have the best results with 220, 380 and 560 features, respectively). **b** The number of features for Db4 wavelet, (*T* test, Entropy and Bhattacharyya distance have the best results with 260, 400 and 600 features, respectively). **c** The number of features for Haar wavelet (*T* test, entropy and Bhattacharyya distance have the best results with 240, 300 and 500 features, respectively). Classification is done by support vector machine with one against one model, RBF kernel and SMO solver

Table 2 Criterion average performance (CAP), wavelet average performance (WAP), classification is done by support vector machine with one-against-one model, RBF kernel and SMO solver

	Sym2	Db4	Haar	CAP (%)
T test	70.03 % (±0.044)	67.38 % (±0.046)	69.19 % (±0.030)	68.86
Entropy	66.78 % (±0.047)	64.35 % (±0.082)	67.23 % (±0.113)	66.12
Bhattacharyya	64.46 % (±0.054)	59.92 % (±0.100)	65.39 % (±0.102)	63.25
WAP	67.09 %	63.88 %	67.27 %	

The values in the parentheses show the standard errors of the means

p value < 0.01) on performance changes depending on the type of wavelet. To elaborate the interaction between feature selection type and wavelet type, the results of Table 2 exhibit that the *T* test criterion has the best performance with all three types of wavelet. Moreover it can be seen that the *T* test criterion works better with the Sym-2 than the other wavelets. In contrary, the Entropy and the Bhattacharyya interact better with the Haar than other wavelets.

Confusion Matrix

Confusion matrix (CM) was computed for the subjects to assess the twelve categories' separability. If a classification system is trained to distinguish between categories, a confusion matrix will summarize the results of testing the algorithm for further inspection. Each column of the matrix represents the instances in a predicted class, while each row represents the instances in an actual class. The name stems from the fact that it makes it easy to see if the system is confusing between classes. A confusion matrix with Sym2 wavelet and *T* test criteria was calculated for all 8 subjects, and results were averaged to create the overall confusion matrix. Classification was performed with SVM and obtained 70 % average performance for all categories, which varied from 41 to 97.8 % (for Sym2 wavelet features).

Table 3 indicates overall averaged values. All values were obtained by *T* test besides Haar, Sym2, and Db4. It is obvious in Table 3 that animals, stationery, and food classes have maximum classification performances, 95.3, 82.3 and 79.3 % accuracies, respectively with the Haar wavelet. Moreover, the same classes have best results for the Sym2 wavelet. On the other hand, dolls, buildings, fruit, and clothes classes with the rate of 51.7, 54.3, 55.1 and 55.7 %, respectively, have minimum classification performances with the Haar wavelet. The same classes also have the worst results with the Sym2 wavelet. To detect categories confusion with each other that led to minimum performances, full complete confusion matrix with *T* test and Sym2 wavelet was presented in Fig. 5.

The confusion matrix was computed for all categories indicating mistakes among the classes. This matrix shows for example, the dolls category is mostly confused with clothing. There are three hypotheses for this problem. At first, maybe the feature reduction method is not optimal, and some useful features of the dolls category might not be chosen, or some common features between dolls and clothing might be mixed, and so some effective features might have been missed. Secondly, some similar colors and shapes in pictures were coded in brain signals, which might not be separated easily by the classifier. Thirdly, some particular groups such as dolls and clothing or fruit and food were confused together because of some similarities among them. It means that fruit is also food, as well as dolls wear clothes. Besides the three discussed hypothesis, results of same categories order for all participants, to some extent, could be dependent on task order.

Informative Features

Extracted features based on training data were counted to introduce informative features. As a result, T5, T4, O2, O1, and C4 were announced as informative channels (O1:23 %, O2:21 %, T4:13 %, T5:11 % C4:10 %). This does not mean that other features had no role in the classification; however, the mentioned features had high role in the categorization task. In addition, D1, D2 and D3 with more than 27, 25 and 21 % repetition were designated as informative DWT scales, respectively. D1, D2 and D3 are the high frequency scales of 1–30 Hz range. For some further clarification, Fig. 6 illustrates EEG wavelet spectrograms of two examples of well-classified categories (animals and stationery). It can be clearly seen from the figures that the most active areas are high frequencies. In the animal category, alpha band frequency shows high energy amplitude In contrast, the stationery category has lower alpha activity and high energy amplitude in beta band frequency.

Table 3 Overall performance for all categories by *T* test criterion with Haar wavelet (first column), Sym2 wavelet (second column) and Db4 wavelet (third column), Classification is done by support vector machine with one against one model, RBF kernel and SMO solver

Categories	Haar (%)	Sym-2 (%)	Db4 (%)
Animal	95.3	97.8	93.3
Flower	74.4	77.9	74
Fruit	55.1	48.7	47
Transportation	70	65	64.9
Body organ	69.8	79	75
Clothes	55.7	57.6	55
Food	79.3	79.3	78
Stationery	82.3	84.6	77.9
Building	54.3	58.8	52.1
Electronic device	66.1	75.4	69
Doll	51.7	41	47.7
Jewel	76.3	75.2	75.1

Study Limitations and Future Work

The performance of SVMs largely depends on the choice of kernels, but the choice of kernel functions, which are well suited to the specific problem, is very difficult. Speed and size are other problems of SVMs both in training and testing. In terms of running time, SVMs are slower than other neural networks for a similar generalization performance. For further studies, other classification methods including artificial neural networks can be examined. To determine the most-involved electrodes and frequency bands of brain signals, a combination of ERP and classification could be used to obtain higher and faster classification performances.

Another limitation is participants, which in such visual experiments are asked to sit still, suppressing or minimizing natural eye and head movements, waiting for and

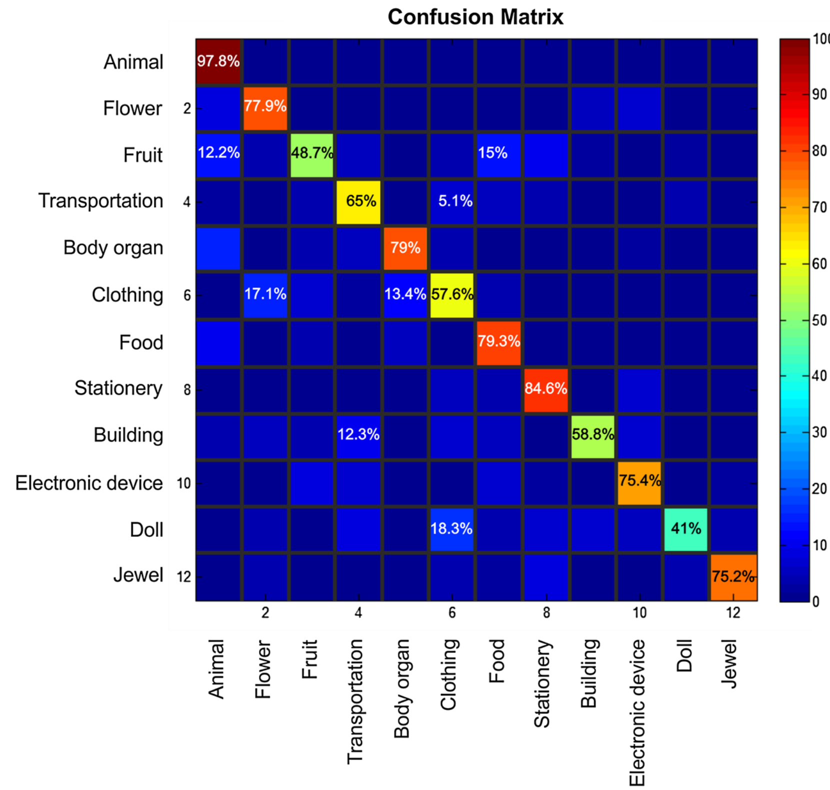

Fig. 5 Confusion matrix. Shows classification accuracies obtained when classifying 12 categories (2D confusion matrix for *T* test and sym2 wavelet) by support vector machine with one against one model, RBF kernel and SMO solver. Fruit category is confused mostly with food and animal classes with 15 and 12.2 %, respectively. Transportation category is confused mostly with clothing (5.1 %). Clothing category is confused mostly with body organs and flower classes with 13.4 and 17.1 %, respectively. Building category is confused mostly with transportation (12.3 %). Doll category is confused mostly with clothing (18.3 %)

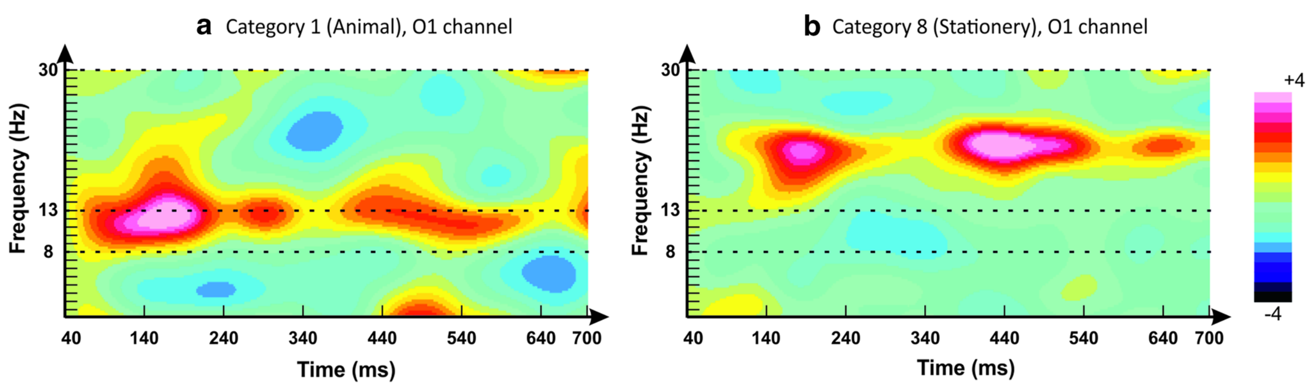

Fig. 6 EEG wavelet spectrogram. Two examples of spectrogram obtained over a representative cortical channel (O1), for one subject and two well-classified categories (animal and stationery). The *dashed horizontal lines* show the borders between the brain frequency bands **a** illustrates EEG alpha activity (813 Hz). In comparison **b** shows EEG beta activity (13–30 Hz) during 140–540 ms

gazing at stimuli because of perceived difficulty of separating brain EEG data from non-brain artifacts. Neuroscience studies often assume that brain activity measured in well-controlled conditions and environments reflects a general principle of brain dynamics during cognitive processing in naturalistic environments. However, until recently, only scattered studies explicitly investigated whether the brain switches to a different method of operation while humans actively behave, adapt to, and interact with ever-changing environments. A BCI communication system with wireless electrodes or improved artifact removing method can be designed for premier applications such as decoding brain signal system for speechless people.

Extra future work can be introduced by optimizing our results. Different feature extraction and feature reduction methods could be applied to raw data, and new features could be checked. However, from the study point of view, distinctive features are the most important issue. In addition, receiver operator characteristic may be a useful method for presenting final results.

Conclusions

The EEG signals of eight participants, during the performance of visual task selecting target images through 12 categories. This study focused on preprocessing and processing units. Statistical ICA and filtering methods were implemented to remove eye-movement and eye blinks. Data were separated into training, validation, and test sets. SVM parameters were setup by “leave one out cross validation” with no interference of test data. Features were extracted using three wavelets (Sym2, Db4, and Haar), and the Haar wavelet yielded the best result with 67.27 % averaged performance. Moreover, features were ranked and reduced by three criteria of scalar feature selection methods namely *T* test, Entropy, and Bhattacharyya distance. *T* test had better outcome among other criteria with 68.86 % averaged performance. Animals, stationery, and food classes have maximum performances of 97.8, 84.6 and 79.3 %, respectively (with Sym2-*T* test combination). Finally T5, T4, O2, O1, and C4 electrodes with D1, D2, and D3 scales were introduced as informative features based on training data.

Acknowledgment The authors thank PAARAND specialized center for (www.paarand.org) their valuable comments and providing the EEG system.

References

Adorni R, Proverbio AM (2009) New insights into name category-related effects: is the age of acquisition a possible factor? Behav Brain Funct 5:33

Bottou L, Chapelle O, DeCoste D, Weston J (2007) Support vector machine solvers large scale kernel machines. MIT Press, Cambridge

Chang CY, Chen SJ, Tsai MF (2010) Application of support–vector–machine-based method for feature selection and classification of thyroid nodules in ultrasound images. Pattern Recognit 43:3494–3506

Choi E, Lee C (2003) Feature extraction based on the Bhattacharyya distance. Pattern Recognit 36:1703–1709

Coyle D, Prasad G, McGinnity TM (2005) A time-series prediction approaches for feature extraction in a brain–computer interface. IEEE Trans Neural Syst Rehabil Eng 13:461–467

Coyle D, McGinnity TM, Prasad G (2006a) Creating a nonparametric brain–computer interface with neural time-series prediction preprocessing. In: The 28th international IEEE engineering in medicine and biology conference, pp 2183–2186

Coyle D, Prasad G, McGinnity TM (2006b) Enhancing autonomy and computational efficiency of the self-organizing fuzzy neural network for a braincomputer interface. In: Proceedings of the FUZZ-IEEE international world congress on computational intelligence, pp 10485–10492

Delorme A, Sejnowski T, Makeig S (2007) Enhanced detection of artifacts in EEG data using higher order statistics and independent component analysis. NeuroImage 34:1443–1449

Demiralp T, Ademoglu A, Schürmann M, Eroglu CB, Başar E (1999) Detection of P300 waves in single trials by the wavelet transform. Brain Lang 66(1):108–128

Fuggetta G, Rizzo S, Pobric G, Lavidor M, Walsh V (2009) Functional representation of living and nonliving domains across the cerebral hemispheres: a combined event-related potential/transracial magnetic stimulation study. J Cogn Neurosci 21:403–414

Garrett D, Peterson DA, Anderson CW, Thaut MH (2003) Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. IEEE Trans Neural Syst Rehabil Eng 11:141–144

Gu Y, Farina D, Murguialday AR, Dremstrup K, Montoya P, Birbaumer N (2009) Offline identification of imagined speed of wrist movements in paralyzed ALS patients from single-trial EEG. Front Neurosci 3:Article No. 62

Higashi H, Tanaka T, Funase A (2009) Classification of single trial EEG during imagined hand movement by rhythmic component extraction. IEEE Int Conf 987:2482–2485

Hoenig K, Sim EJ, Bochev V, Herrnberger B, Kiefer M (2008) Conceptual flexibility in the human brain: dynamic recruitment of semantic maps from visual, motor, and motion-related areas. J Cogn Neurosci 20:1799–1814

Hsu CW, Chang CC, Lin CJ (2010) A practical guide to support vector classification. Bioinformatics 1:1–16

Hsu C-W, Lin C-J (2002) A comparison of methods for multiclass support vector machines. IEEE Trans Neural Netw 13(2):415–425

Ince NF, Tewfik A, Arica S (2005) Classification of movement EEG with local discriminant bases. In: IEEE international conference on acoustics, speech and signal processing, pp 414–416

Johnson JS, Olshausen BA (2003) Time course of neural signatures of object recognition. J Vis 3:499–512

Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui V, Sejnowski TJ (2000) Removing electroencephalographic artifacts by blind source separation. Psychophysiology 37:163–178

Kandaswamy A, Kumar CS, Ramanathan RP, Jayaraman S, Malmurugan N (2004) Neural classification of lung sounds using wavelet coefficients. Comput Biol Med 34(6):523–537

Keshtiban AM, Razmi H, Kozehkonan MK (2011) Combined LVQ neural network and multivariate statistical method employing wavelet coefficient for EEG signal classification. In: IEEE international conference on mechatronics (ICM), pp 809–814

Kiefer M (2001) Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. Mem Cognit 29:100–116

Kressel U (1999) Pairwise classification and support vector machines. In: Schölkopf B, Burges C, Smola A (eds) Advances in kernel methods—support vector learning. MIT Press, Cambridge, pp 255–268

Krishnan Mookiah MR, Rajendra Acharya U, Lim CM, Petznick A, Suri JS (2012) Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features. Knowl Based Syst 33:73–82

Lal TN, Schroder M, Hinterberger T, Weston J, Bogdan M, Birbaumer N, Schlkopf B (2004) Support vector channel selection in BCI. IEEE Trans Biomed Eng 51:1003–1010

Martinovic J, Gruber T, Muller MM (2008) Coding of visual object features and feature conjunctions in the human brain. PLoS ONE 3(11):e3781. doi:10.1371/journal.pone.0003781

Martinovic J, Mordal J, Wuerger S (2011) Event-related potentials reveal an early advantage for luminance contours in the processing of objects. J Vis 11:1–15

Merry RJE (2005) Wavelet theory and applications: a literature study. Technische Universiteit Eindhoven, Eindhoven

Messer SR, Agzarian J, Abbott D (2001) Optimal wavelet denoising for phonocardiograms. Microelectron J 32(12):931–941

Milgram J, Cheriet M, Sabourin R (2006) “One against one” or “one against all”: which one is better for handwriting recognition with

SVMs? In: Tenth international workshop on frontiers in handwriting recognition: inria-00103955, version 1
Miller GA, Lutzenberger W, Elbert T (1991) The linked-reference issue in EEG and ERP recording. J Psychophysiol 5:279–280
Misaki M, Kim Y, Bandettini PA, Kriegeskorte N (2010) Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. NeuroImage 53:103–118
Muller KR, Tangermann M, Dornhege G, Krauledat M, Curio G, Blankertz B (2008) Machine learning for real-time single-trial EEG-analysis: from brain–computer interfacing to mental state monitoring. J Neurosci Methods 167:82–90
Palaniappan R, Paramesran R, Nishida S, Saiwaki N (2002) A new brain–computer interface design using fuzzy ARTMAP. IEEE Trans Neural Syst Rehabil Eng 10:140–142
Paz-Caballero D, Cuetos F, Dobarro A (2006) Electrophysiological evidence for a natural/artificial dissociation. Brain Res 1067:189–200
Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. NeuroImage 45:199–209
Peters BO, Pfurtscheller G, Flyvbjerg H (2001) Automatic differentiation of multichannel EEG signals. IEEE Trans Biomed Eng 48:111–116
Philiastides MG, Sajda P (2005) Temporal characterization of the neural correlates of perceptual decision making in the human brain. Cereb Cortex 16:509–518
Phillips S, Takeda Y, Singh A (2012) Visual feature integration indicated by phase-locked frontal-parietal EEG signals. PLoS ONE 7(3):e32502. doi:10.1371/journal.pone.0032502
Pregenzer M, Pfurtscheller G (1999) Frequency component selection for an EEG-based brain to computer interface. IEEE Trans Rehabil Eng 7:413–419
Proverbio AM, Del Zotto M, Zani A (2007) The emergence of semantic categorization in early visual processing: ERP indices of animal vs. artifact recognition. BMC Neurosci 8:8–24
Pulvermuller F, Lutzenberger W, Preissl H (1999) Nouns and verbs in the intact: evidence from event-related potentials and high-frequency cortical responses. Cereb Cortex 9:497–506
Rafiee J, Rafiee MA, Prause N, Schoen MP (2011) Wavelet basis functions in biomedical signal processing. Expert Syst Appl 38:6190–6201
Sanei S, Chambers JA (2007) EEG signal processing. Wiley, New York. ISBN-10: 0470025816
Sherwood J, Derakhshani R (2009) On classifiability of wavelet features for EEG-based brain–computer interfaces. In: International joint conference on neural networks (IJCNN), pp 2895–2902
Simanova I, Van Gerven M, Oostenveld R, Hagoort P (2010) Identifying object categories from ERP toward decoding of conceptual representations. PLoS ONE 5(12):e14465. doi:10.1371/journalpone0014465
Subasi A (2005) Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. Expert Syst Appl 28:701–711
Sykacek P, Roberts S, Stokes M, Curran E, Gibbs M, Pickup L (2003) Probabilistic methods in BCI research. IEEE Trans Neural Syst Rehabil Eng 11:192–195
Tzovara A, Murray MM, Plomp G, Herzog MH, Michel CM, DeLucia M (2012) Decoding stimulus-related information from single-trial EEG responses based on voltage topographies. Pattern Recognit 45:2109–2122
Vapnik VN (1998) Statistical learning theory, 1st edn. Wiley, New York
Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM (2002) Brain–computer interfaces for communication and control. J Clin Neurophysiol 113:767–791
Zhang A, Yang B, Huang L (2008) Feature extraction of EEG signal using power spectral entropy. IEEE Int Conf Biomed Inform 978:435–438