Content based Image Retrieval Databases Classification with Brain Event Related Potential

Rodrigo Prior Bechelli
Centro Universitário da FEI, Av. Humberto de A. C. Branco, 3972, 09850-901, São Bernardo do Campo, São Paulo, Brazil

Keywords: Brain Computer Interface, BCI, Content Based Image Retrieval, CBIR, Electroencephalography, EEG.

Abstract: This paper evaluates and compile information related to Electroencephalography (EEG) used as a pattern to classify a Content Based Image Retrieval (CBIR) system based on an Event Related Potential (ERP) as an input data vector to classify an image database. The Rapid Serial Visual Presentation (RSVP) is used as a method to present multiple images to obtain a series of P300 brain response and specify the duality of target or non- target images (oddball paradigm).

1 INTRODUCTION

The research field of Content Based Image Retrieval (CBIR) proposes methods to search across a high volume of image or data files. Over the last few decades it has increased exponentially (Liu et al., 2007) and even with the most advanced technology we have still not bridged the gaps between what we search and the conceptual idea behind what we are trying to search (Deserno et al., 2009).

One initiative to connect these gaps is to understand how we process these ideas using acquisition methods based on brain activity to possibly achieve better query results. Based on that proposition, this work explores a Brain Computer Interface (BCI) based on brain Event Related Potential (ERP) (Farwell and Donchin, 1988) using $P300$ component captured via Electroencephalography (EEG) signals applied recently by authors over CBIR queries.

2 OBJECTIVES

- Evaluate the actual classification methods of applying EEG signals to an image database;
- Evaluate and compare the methods of feature vector extraction from EEG signals to apply in an image database;
- Query an image database classified via EEG signals and evaluate its results with a CBIR methodology.

3 STATE OF THE ART

The following authors, Wang et al. (2009); Khosla et al. (2011); Ušćumlić et al. (2011); Healy and Smeaton (2011); Ušćumlić et al. (2013); Mohedano et al. (2014) were compared to map the state of the art of applying EEG over CBIR and specifically define a path of research based on: *acquisition hardware, Rapid Serial Visual Presentation (RSVP), image database, tested subjects, preprocessing, processing and classification*. With this map it is possible to evaluate, in a future work, all the requirements to create a simulation environment.

To capture the brain signals, EEG acquisition was applied and all researched authors adopted the 10-20 protocol to distribute the electrodes on the scalp. Also, all the authors proposed the method of Event Related Potential (ERP) (Sutton et al., 1965) to trigger the $P300$ brain response and specify the duality of target or non-target images (*oddball paradigm*) (Donchin et al., 1978; Donchin and Arbel, 2009).

3.1 Acquisition Hardware and Environment

The acquisition hardware details and environment to capture the EEG signals, for example a *Faraday cage* environment, are not described in all references, but the ones that exposes it are detailed in Table 1.

Bechelli, R.
Content based Image Retrieval Databases Classification with Brain Event Related Potential.
In *Doctoral Consortium (DCBIOSTEC 2016)*, pages 3-8

Table 1: Hardware and Environment.

Hardware
64-channel BioSemi ActiveTwo system in an extended 10-20 montage at a sampling rate of 2048Hz (Wang et al., 2009)
no description (Khosla et al., 2011)
no description (Ušćumlić et al., 2011)
16 channel KT88-1016 EEG system with a left mastoid reference and the chin as ground. Ag/AgCl electrodes were used with a 10-20 placement. Digitized at 100Hz and subsequently band-passed from 0.1Hz to 20Hz (Healy and Smeaton, 2011)
64-channel BioSemi ActiveTwo system in an extended 10-20 montage at a sampling rate of 2048Hz. Calibrated via electrooculographic (EOG) activity (Ušćumlić et al., 2013)
31 channel BCI with a sample rate of 1kHz was used to capture the brain reaction of the users during the image presentation. The electrodes were located according to the 10-20 system distribution (in a Faraday Cage) (Mohedano et al., 2014)

3.2 Rapid Serial Visual Presentation (RSVP)

To evaluate several images for each one of the subjects, the Rapid Serial Visual Presentation (RSVP) was used. This method consist in presenting several images to the subject in a controlled environment and frequency such as in a flash mode. The sequence of flashes contains several images while it is required to the subject to identify a target in a series of non target images (Mohedano et al., 2014).

Over the series of articles researched the same method is applied with some variation as presented in Table 2.

Khosla et al. (2011) described a possible issue on dissimilar image sequences that have different colors, scales and textures. These differences could trigger a ”false alarm” $P300$ signal based on the surprise. They also propose a preprocessing method of feature extraction and a distance model between images to evaluate the image database.

3.3 Image Database

In order to compare the results in CBIR environment a image database must be used to evaluate the resulting queries. As long as there are trusted image databases available for research the studied papers are compared by this definition in Table 3.

Table 2: Rapid Serial Visual Presentation (RSVP).

RSVP Method Details
240px x 240px image size, 62 objects database randomly partitioned in two, with 1000 images as 10 blocks of 100 images each presented at 6Hz (Wang et al., 2009)
256px x 256px image size, presented at 10Hz and results presented in 50 images screen with 10 images/row (Khosla et al., 2011)
220 image sequences with 20 target images presented at 4Hz (Ušćumlić et al., 2011)
Target images were inserted into blocks of 100 non-target presented at 10Hz. Total of 4.800 images with 60 target randomly distributed (Healy and Smeaton, 2011)
20 images per object (eg. ”dog”, ”eagle” etc) with 10 categories. 10% targets. Subjects sat at 60cm from screen with images occupying approximate 6^o x 4^o of their visual field and were instructed to silently count images of a specified object (Ušćumlić et al., 2013)
22 images selected with a single object and limited complexity background. With 192 window of each image presented at 5Hz zoomed and centered at screen. Subjects were instructed to count the number of windows containing a part of the object (Mohedano et al., 2014)

Table 3: Image Database.

Image Databases
Caltech 101 dataset (3798 images) and 62 object from Caltech and Satellite imagery with ”Helipad” 1051 targets (Wang et al., 2009)
no description (Khosla et al., 2011)
Custom database with 1382 images annotated using features based on human vision Colored Pattern Appearance Model (CPAM) and Edge-based features (Ušćumlić et al., 2011)
Amsterdam Library of Object Images (ALOI) with 1000 objects with 4800 images with a number of different camera angles/lighting conditions (Healy and Smeaton, 2011)
Caltech dataset (Ušćumlić et al., 2013)
Berkeley Segmentation Dataset and Benchmark (BSDB) (Mohedano et al., 2014)

3.4 Tested Subjects

All the articles used primary data collected in sessions of EEG and RSPV following the methods already described. The total number of subjects in each experiment is relevant in order to compare the volume of information captured. The details for each experiment are described in Table 4.

Table 4: Subjects.

Subjects and Details
4 undergraduate and graduate students, staff and faculty that were not digital media analysts, but were familiar with EEG work (Wang et al., 2009)
no description (Khosla et al., 2011)
1 subject for database classification. Retrieval stage was made with synthetic data that emulates EEG decoding with error rates of 30% and 60% (Ušćumlić et al., 2011)
8 postgraduate and staff population on campus, 5 males and 3 females were recruited with an average age of 27.5 years with standard deviation of 4.5 years. (Healy and Smeaton, 2011)
15 subjects with normal or corrected-to-normal vision. There was no specific criteria for recruiting the subjects (Ušćumlić et al., 2013)
5 subjects between 21 and 32 years old (Mohedano et al., 2014)

As detailed by Ušćumlić et al. (2013), the EEG data captured from the subjects in a live demonstration and in a pubic space, presented at a high noise environment. All the other papers are, implicitly, conducted in a laboratory environment.

3.5 Software

None of the articles expose the BCI software platform used to develop the test, acquiring and processing EEG data. This result is expected as long as some of the acquisition hardware already come with its own software packages.

4 PROCESSING DATA

All the researched work employ a method of classification based on EEG data and a method to propagate the extracted features for the remaining target images of the database. In this section it will be pointed the described methods in each evaluated papers and the exploration of each method will be evaluated in future works.

4.1 EEG Processing

Once EEG data is acquired, RSVP must be processed in order to obtain the ERP signals and generate a series of features that can be propagate to the additional unclassified images and detailed in Table 5.

Table 5: EEG Processing.

EEG Processing Algorithms
Apply a method to detect in real time the $P300$ signal using Fisher Linear Discriminator using a window of $100ms$ (Wang et al., 2009)
no description (Khosla et al., 2011)
Gaussian classifier with no additional details (Ušćumlić et al., 2011)
Support Vector Machine (SVM) (Healy and Smeaton, 2011)
Use a two step method applying a Canonical Variate Analysis (CVA) initially and a Gaussian Classifier over the most discriminant features (Ušćumlić et al., 2013)
Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel classifier with normalized feature vector with zero mean and unit standard deviation across each feature component (Mohedano et al., 2014)

Additionally to selected articles in Table 5, Tong and Chang (2001) worked with linear kernel and authors evaluated the effect of a reduced number of EEG channels applying a Sequential Forward Feature Selection (SFFS) and Somol et al. (1999) tried to find subsets between the determined classes of images.

4.2 Image Database Processing

Khosla et al. (2011) implemented some level of processing at the image database. It was highlighted the need to normalize images at color, background and texture of images in order to reduce the false positive levels of $P300$ signals in EEG capture. First is implemented a color conversion from RGB to HSV and implement Euclidean distance calculation over image histogram. Once calculated images are ordered and clustered.

The state of the art compared six articles that apply the same technique: ERP, $P300$, RSVP and CBIR. Even with the same basic structure and dealing with the same challenges these papers presented a level of variance in the features that were evaluated. But it also possible to detect some level of concordance in the image database: Caltech image database (Wang et al., 2009; Uščumlić et al., 2013). In RSVP it is not possible to have the same description in all articles but there is possible to verify a trend in image size and frequency rates. In EEG Processing it is also possible to identify to use of Support Vector Machine (SVM) (Healy and Smeaton, 2011; Mohedano et al., 2014) and Gaussian Classifier (Uščumlić et al., 2011, 2013).

5 METHODOLOGY AND EXPECTED OUTCOME

A RSVP protocol test were develop to use *Caltech 101 dataset* (Fei-Fei et al., 2007) to be preprocessed as described in section 3. This database have 9.144 images, classified in 102 categories. For this test 57 categories were adopted with total of 3.436 images. Images were preprocessed to fit 300*px* height and 300*px* width.

The RSVP test were developed to work on a computer screen of 1024*x*768*px* resolution (but also work with some variation of this resolution). It starts with 5*s* of black image presentation and after this time the images were centered in a black background and presented sequentially at 4*Hz* frequency.

In this stage, the consumer product Emotiv-EPOC+ (Emotiv, 2011) was used to capture EEG signals. This device is composed of 14 electrodes ($AF3$, $F7$, $F3$, $FC5$, $T7$, $P7$, $O1$, $O2$, $P8$, $T8$, $FC6$, $F4$, $F8$, $AF4$), with acquisition rate of 128*Hz*. It is known that this device partially cover occipital and parietal electrodes, important regions to acquire Visual Evoked Potential (VEP) as previously reported by Duffy et al. (1999). But recent research successfully employed the Emotiv-EPOC+ in order to capture P300 responses as proposed by Ekanayake (2010).

A build script was developed with Python (Oliphant, 2007; Pérez and Granger, 2007; Gramfort et al., 2013) program language to interface a test RSVP application with OpenViBE (Renard et al., 2010) platform.

The test process flow described as follow:

1. Subject basic information was obtained during an initial interview to verify any medical or physiological restriction;
2. Subject read and sign an informed consent;
3. Subject is prepared for EEG acquisition equipment;
4. Test is built specifically for the subject following expected requirements:
   (a) Subject Name;
   (b) Subject Born Date;
   (c) Subject Gender;
   (d) Image category choose for test (e.g: ”dog”);
   (e) Number of targets in RSVP test (default value 30);
   (f) Minimum number of non-targets between each target (default value 6);
   (g) Maximum number of non-targets between each target (default value 20);
   (h) Total number of sessions per subject (default value 4);
   (i) RSVP Frequency Rate in *Hz* (default value 4).
5. The test consist in 4 sessions of 2*min* with a range of 400 to 600 images with 30 target images;

With this information it is expected to use a classification method for the acquired feature vectors of target and no-target images in order to tag the image database.

Once the image database is tagged it will be possible to evaluate results achieved using a BCI classification system against a manual classification of a CBIR.

6 STAGE OF THE RESEARCH

The python software that build the RSVP protocol integrated with OpenViBE were developed an tested with a subset of *Caltech 101 dataset*.

In order to validate the entire pipeline, EEG data were acquired using Emotiv- EPOC+ from 6 undergraduate and graduate students volunteers between 22 and 40 years old. All subjects have a certain level of knowledge of the tests, but none of them work directly with this research. Tests were conducted in the university usability laboratory.

At this stage, the analysis of the collected data was started. In Figure 1 is possible to evaluate the average of EEG signal of the evoked potentials from a single subject of 120 presented target images along all four sessions.

Evaluating the results in Figure 1, the P300 signal is expected to be displayed in time window from 250*ms* to 450*ms*, showing expected responses for this scenario (Duvinage et al., 2012).

From the same subject, in Figure 2, it is possible to identify the expected activation areas at 300*ms* in electrodes $O1$, $O2$, $P8$ and $P7$.

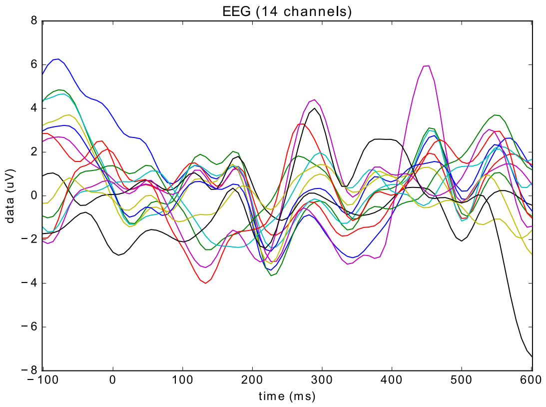

Figure 1: P300 analysis from a single subject. Average from 120 target images (30 target images from all sessions).

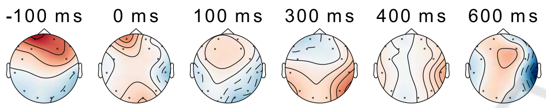

Figure 2: Topographic map of evoked potentials from Figure 1.

None of this data was cleaned from additional artifacts, for example, eye movement electromyography signals or invalidation of time window epochs. At this stage it necessary to evaluate a preprocessing algorithm that better suit the research purpose.

ACKNOWLEDGEMENTS

This study was produced with FEI, CAPES and FAPESP funding.

REFERENCES

Deserno, T. M., Antani, S., and Long, R. (2009). Ontology of gaps in content-based image retrieval. *Journal of digital imaging*, 22(2):202–15.

Donchin, E. and Arbel, Y. (2009). P300 based brain computer interfaces: A progress report. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5638 LNAI, pages 724–731.

Donchin, E., Ritter, W., and McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP.

Duffy, F. H., Iyer, V. G., and Surwillo, W. W. (1999). *Eletroencefalografia Clínica e Mapeamento Cerebral Topográfico: Tecnologia e Prática*. Revinter Ltda, 1 edition.

Duvinage, M., Castermans, T., Dutoit, T., Petieau, M., Hoellinger, T., Saedeleer, C., Seetharaman, K., and Cheron, G. (2012). A P300-based quantitative comparison between the emotiv epoc headset and a medical EEG device. *Proceedings of the 9th IASTED International Conference on Biomedical Engineering, BioMed 2012*, pages 37–42.

Ekanayake, H. (2010). P300 and Emotiv EPOC: Does Emotiv EPOC capture real EEG? *Web publication http://neurofeedback. visaduma. info/ . . .* , page 16.

Emotiv, S. (2011). EMOTIV EPOC: Brain Computer Interface & Scientific Contextual EEG.

Farwell, L. a. and Donchin, E. (1988). Talking Off the Top of Your Head. *electroencephalography and clinical Neurophysiology*, 70(6):510–523.

Fei-Fei, L., Fergus, R., and Perona, P. (2007). Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1):59–70.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., and Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, (7 DEC).

Healy, G. and Smeaton, A. (2011). Optimising the number of channels in EEG-augmented image search. *Proceedings of the 25th BCS Conference on . . .* , pages 1–6.

Khosla, D., Bhattacharyya, R., Tasinga, P., and Huber, D. J. (2011). Optimal Detection of Objects in Images and Videos Using Electroencephalography (EEG) Deepak. In Kadar, I., editor, *SPIE Defense, Security, and Sensing*, pages 80501C–80501C–11. International Society for Optics and Photonics.

Liu, Y., Zhang, D., Lu, G., and Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40:262–282.

Mohedano, E., Healy, G., McGuinness, K., Giró-i Nieto, X., O'Connor, N. E., and Smeaton, A. F. (2014). Object Segmentation in Images using EEG Signals. In *Proceedings of the ACM International Conference on Multimedia - MM '14*, pages 417–426, New York, New York, USA. ACM Press.

Oliphant, T. E. (2007). SciPy: Open source scientific tools for Python. *Computing in Science and Engineering*, 9:10–20.

Pérez, F. and Granger, B. E. (2007). IPython: A system for interactive scientific computing. *Computing in Science and Engineering*, 9(3):21–29.

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., and Lécuyer, A. (2010). OpenViBE: An Open-Source Software Platform to Design, Test, and Use Brain–Computer Interfaces in Real and Virtual Environments.

Somol, P., Pudil, P., Novovičová, J., and Paclík, P. (1999). Adaptive floating search methods in feature selection. *Pattern Recognition Letters*, 20(11-13):1157–1163.

Sutton, S., Braren, M., Zubin, J., and John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, 150(700):1187–1188.

Tong, S. and Chang, E. (2001). Support vector machine active learning for image retrieval. *Proceedings of the ninth ACM international conference on Multimedia MULTIMEDIA 01*, 54(C):107.

Ušćumlić, M., Chavarriaga, R., and Millán, J. d. R. (2011). On coupling Computer Vision and Single trial detection of ERP under RSVP. *International Journal of Bioelectromagnetism*, 13(3):133–135.

Ušćumlić, M., Chavarriaga, R., and Millán, J. D. R. (2013). An iterative framework for EEG-based image search: robust retrieval with weak classifiers. *PloS one*, 8(8):e72018.

Wang, J., Pohlmeyer, E., Hanna, B., Jiang, Y., Sajda, P., Chang, S., and Jun Wang, Eric Pohlmeyer, Barbara Hanna, Yu-gang Jiang, Paul Sajda, S.-f. C. (2009). Brain State Decoding for Rapid Image Retrieval. In *Proceedings of the seventeen ACM international conference on Multimedia*, pages 945–954.