Investigating EEG for Saliency and Segmentation Applications in Image Processing

Eva Mohedano Robles, August 2013

***Abstract*—The main objective of this project is to implement a new way to compute saliency maps and to locate an object in an image by using a brain-computer interface. To achieve this, the project is centered in designing the proper way to display the different parts of the images to the users in such a way that they generate measurable reactions. Once an image window is shown, the objective is to compute a score based on the EEG activity and compare its result with the current automatic methods to estimate saliency maps. Also, the aim of this work is to use the EEG map as a seed for another segmentation algorithm that will extract the object from the background in an image. This study provides evidence that BCI are useful to find the location of the objects in a simple images via straightforward EEG analysis and this represents the starting point to locate objects in more complex images.**

***Index Terms*— Brain-computer interfaces (BCI), Eelectroencephalography (EEG), segmentation, saliency map, rapid serial visual presentation (RSVP)**

I. MOTIVATION

THIS paper presents a final report on a project in the School of Electronic Engineering carried out as part of an Erasmus program in the research center CLARITY at DCU. The work consists of exploring what information it is possible to extract from a Brain Computer Interface (BCI) during a local exploration of an image containing an object of interest. Specifically, the project has a double purpose: on one hand, the work is focused on finding out if BCI devices are useful to estimate visual saliency maps. On the other hand, the aim is to use EEG signals to extract the location of an object in the whole image and perform its segmentation.

Visual saliency maps automatically estimate which regions in the image mostly attract the attention of the user. They are computed by the intrinsic features of the image such as color, texture, orientation, intensity, etc. The purpose of this work is to compute these kind of maps based directly on the brain response of the user, instead of applying computer vision techniques on the image. A similar purpose has been realized in other works with eye tracker devices [1], and it has shown a correlation between the saliency maps and the inspection of the image by the user. However, BCI devices have not been used to compute these maps to the best of the author's knowledge.

Concerning the segmentation, the system would be a new interaction mechanism to segment an image, where the "interaction" would be reduced to the minimum expression: the user is just asked to look at the presentation of different image blocks. This way, s/he would be released of any kind of manual task like drawing a box around the object of interest or drawing scribbles on the object and the background [2]. In this work, the semi-supervised segmentation algorithm will be seeded directly by the reaction of the brain.

II. RELATED WORK

Previous works combining Brain Computer Interfaces (BCI) and computer vision [3][4][5] have been mainly focused in image retrieval and object detection. In these works, the way to present the images follows the *oddball paradigm*. This approach consists in presenting a "target" image between a large amount of "distractor" images in a Rapid Serial Visual Presentation (RSVP). The images are presented at a high rate, around 10Hz, in such a way that a specific signature in their EEG signals is produced when the user sees the target images (or rare stimulus). This signature is known as P300 wave and it is a kind of Event-Related Potential (ERP) related to the process of the recognition of a specific visual stimulus. The wave consists mainly in a positive peak in the EEG wave after 300ms following the visual stimulus.

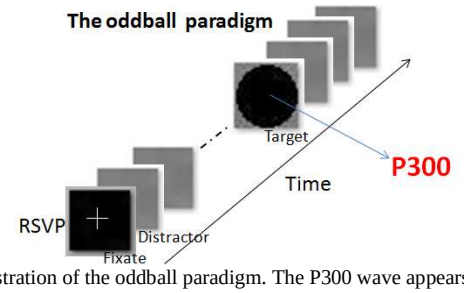


Fig. 1 Ilustration of the oddball paradigm. The P300 wave appears in the EEG data acquired 300ms after the target image.

Two previous works of a BCI system applied to image retrieval and detection were presented by Wang (2009) [3] and Healy (2011) [4]. In both cases the authors perform a RSVP at 10 Hz of images from known datasets (Caltech and ALOI, respectively) to detect those images in which a specific object appears. It is remarkable that in Wang's paper the user is not asked to press any additional button when a target image is seen.

The main reference for this project was from Bigdely-Shamlo (2008) [5], where satellite images are explored by local windows to detect which of them contain airplanes. Nevertheless, such work differs from the goal of this work, where objects may be distributed in multiple adjacent windows. This project targets a challenging approach because

it focuses on target windows instead of target images. This means that the object of interest may be partially included in a window. It is possible that the size ratio between the object part and the window will influence the associated EEG response.

III. LOCAL EXPLORATION OF THE IMAGE

A. Input EEG Device.

The EEG device used in this project is the KT88-1016, the same one used in [4]. The sampling rate is 100Hz and offers 16 channels of acquisition. Both of these features are low resolution compared to other studies [3][5], which indicates that this research exploits basic low-cost acquisition equipment. In addition, it was decided to adopt an even simpler configuration, by just considering the 8 channels located mainly at the bottom of the head. These channels were chosen because this area is the most sensitive to P300 detection.

B. Sliding Window Interface

A first version of the interface to present the images was developed from scratch in Python. It consisted in a black mask that covered the entire image except one square region. The presentation was the movement of this window across the image with a continued scan: the window started in the top-left corner of the image and moved to the right. When it arrived to the border, it moved down the size of the window, and started the inspection to the left. This movement was repeated until all the regions of the image were shown.

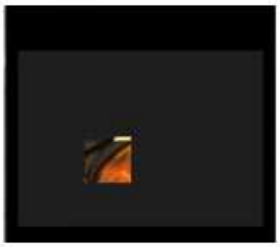


Fig 2 Screen shot of the Sliding window interface.

The time reference of the local computer and the position of the window in the image is recorded during the presentation. Simultaneously the EEG activity of 8 channels is recorded relative to the local computer time, so it is possible to associate the brain reaction of the user to each position in the window.

C. Problems of the Sliding Window

After running the first experiment with one user and obtaining only noise, we analyzed the possible problems with this implementation. We held a videocall with two experts in the neureosicence field: Thomas Ward and Nima Bidgely Shamlo, one of the authors of [5], from the Center for Computational Neuroscience of San Diego. After the discussion, the following drawbacks were identified:

1- The movement of the window around the screen forced the users to **move the eyes** during the presentation and it is known that this generates artifacts on the acquired EEG signal.
2- The **progressive exploration** of the image may not generate any useful response in the EEG waves due to the fact that the brain mainly reacts to abrupt changes.
3- Due to the size of the objects of the images selected, the **amount of target windows was too high**, another issue that may hamper the triggering of any useful reaction in the brain.
4- **The way to synchronize** the time of the visual stimulus and the EEG activity may generate misalignements due to possible delays between the script for the presentation and the one for the acquisition.

D. Second Design: Random RSVP at local scale based on the SNAP interface.

In order to fix the problems of the Sliding Window Interface, a second design based in the Simulation and Neuroscience Application Platform (SNAP) developed in the Swartz Center for Computational Neuroscience[1] was adapted to the purpose of this project.

The new implementation consisted in cropping the images in the different windows to later display the windows in a random order following the RSVP approach at 10Hz of frequency.

Fig 3 On the left, windows cropped from one image (The windows containing the flower are considered the target windows, the rest are the distractor windows). On the right, screen shot of the SNAP interface adapted: All the windows of the image are displayed in the same fixed position in a random order.

With the new implementation:

1- The window of the image was in a **fixed position** on the screen, to avoid the movement of the eyes.
2- The inspection of the image was **random** instead of progressive.
3- We built a brand new **controlled image dataset** instead of the Grabcut dataset, previously used due to its popularity in interactive segmentation work. The new dataset consists in 32 natural photos with their manually generated ground truth masks. Each image presents a single salient object in a uniform background. The main feature is that the size of the object is small compared to the size of the

[1] http://sccn.ucsd.edu/

background (around 15% of target window for each image).

4- The best way to fix the problem of the synchronization would have been using one of the channels of the EEG device as a signal to mark when the visual events happen. But due to the extra time that would be required to implement this, it was decided to keep **the same method of synchronization** (same computer time for the visual events and EEG acquisition) for the first trials.

With the changes 1) and 2) the approach became more similar to the oddball paradigm, because a few stimuli were presented between a large amount of distractors. Nevertheless, the rate of targets is still higher if it is compared with the 1% of target images used in [3].

IV. Experimental Set-up

After the review of the interface, the effort was focused in trying to detect some "easy reactions" before running the main experiment. The reason was to make sure that the device was working properly and the synchronization method was good enough to identify the visual stimuli with the brain response from user.

A. Checking the electronics and synchronization:

1) Alpha waves

When the user has closed the eyes the dominant frequency of the brain is around 8Hz-12Hz [7]. These kind of waves are known as alpha waves and they are easy to detect even in the time domain. Visualizing these waves in real time before any experiment provides an easy way to make sure that the EEG device is working properly.

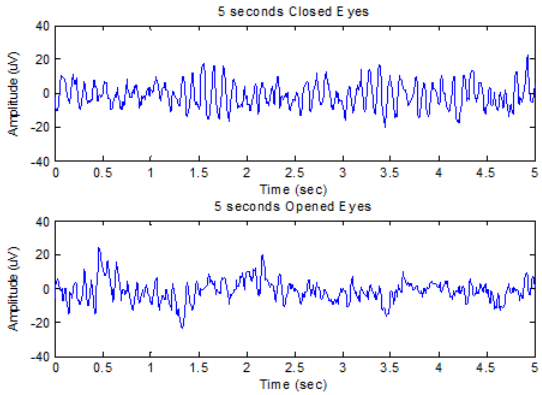


Fig 4. 5 seconds of closed and opened eyes. The waves look different in the time domain, this fact becomes an easy way to check if the device is properly connected.

2) Detecting ERPS from a series of flashes

When the user is exposed to abrupt visual changes, like a white flash after seeing a black screen, a specific ERP is generated. The ERP associated to a flash presents a positive peak 100ms (P100) after the flash stimuli, and a negative peak around the 150-200ms (N100). Related literature and discussions with Dr. Graham Healy and Dr. Michael Keane suggested that a good way to find the ERP wave forms is by averaging a large amount of reactions to the flashes. By averaging, the high noise present in EEG signals is canceled and it is possible to see the ERP.

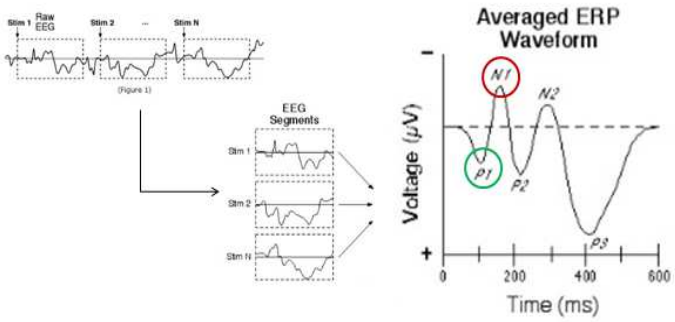
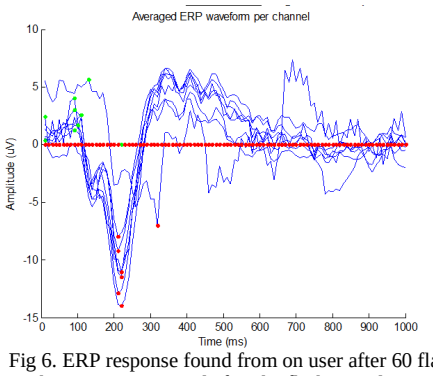


Fig 5. Average technique to find the ERP response[2].

Figure 6 is the average of 60 flashes presented to one user. The experiment consisted in presenting one flash each 2 seconds.

This study ensures correct synchronization of the visual events. This test also indicated that the detection of a specific waveform requires the repetition of the same stimulus several times, 60 times in this experiment.
This fact highlights that it would be probably necessary to display multiple times the different windows of an image in the final experiment to obtain a clear EEG waveform of the brain reaction, at least, in the time domain.



Channel	P100 (ms)	N1 (ms)
1	130	320
2	90	210
3	90	210
4	10	220
5	110	220
6	90	210
7	10	220
8	100	220
Mean	**80**	**230**

Fig 6. ERP response found from on user after 60 flashes. The lines represent the average 1 second after the flash stimuli in each of the 8 considered channels. The first positive peak for each signal is marked in green (P100), and the main negative peak (N100) is marked in red. The exact time values for each channel and their average are also provided in the table near the figure.

3) Simplifying the images to the easiest case: Synthetic images

The results obtained from the first trials were evidence that it is possible to generate some detectable reaction in the brain. The next challenge was to test the presentation scheme described in Section III.D and how to process the captured EEG signals. To reduce the complexity of the experiment, the collection of 32 real images was replaced by 4 synthetic images where a geometric shape is fitted to a window (Fig 7)

[2]Figure extrated from https://uwaterloo.ca

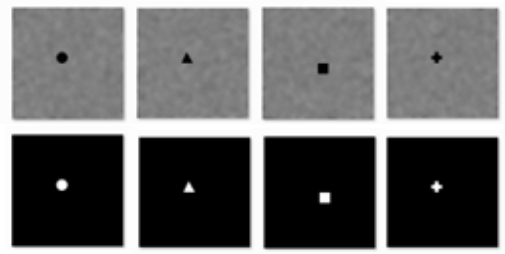


Fig 7. Synthetic image dataset and their ground truth masks.

Each image has a resolution of 300x300 pixels and they were cropped into a 30x30 pixel windows, having an amount of 100 windows per image, where only one window is a target window.

Using these images in the experiment allows inspection of each window more than once. Otherwise there would not be enough target examples to train a classifier because there is only one target among the 100 windows per image. In addition, the results of detecting the ERPS of the flashes is evidence that it is necessary to repeat several times the stimulus to see a clear waveform.

V. Signal Processing of EEG Signals

1) The experiment

The experiment was run in the Faraday Cage of the Nursing Building of Dublin City University. The room is designed specifically to run EEG experiments and isolates the user from external noises.

The 4 synthetic images were displayed by the adapted SNAP interface to one user who was completely free of any mechanical interaction. Each image was presented 32 times and after 8 repetitions of each image (~5 minutes of presentation), the user had a rest period.

2) Data acquired

An amount of 128 images were displayed (32 repetitions for each one of the 4 shapes), having an amount of 32 examples of target windows and 3,168 examples of distractor windows.

3) Data preprocessing

The data was processed with Matlab 7.12. The 8 EEG raw data obtained (one for each connected channel) was low-pass filtered to 50Hz and normalized to 0 mean and standard deviation 1 as suggested in [8].

Each presented window was associated to 1 second of preprocessed EEG activity after its presentation. So, each presented window corresponded to 8 feature vectors of 100 samples corresponding to 1 second after the stimulus presentation for each EEG channel.

4) Single Trial

The process to extract the proper features is critical for successful classification of the windows. The main challenge corresponds to a high variability in the waveforms of target and distractor stimuli [9]. Figure 8 presents two single exemplars from the two classes, while Figure 9 plots the overlap of all considered single trials.

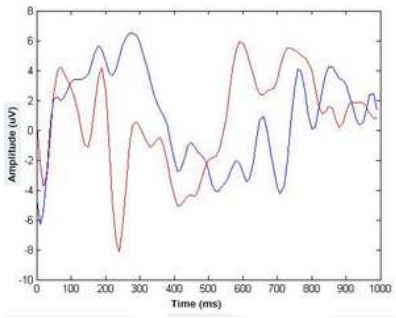


Fig 8. One second of EEG response of one of the channels for one single trial of target window (red) and distractor window (blue).

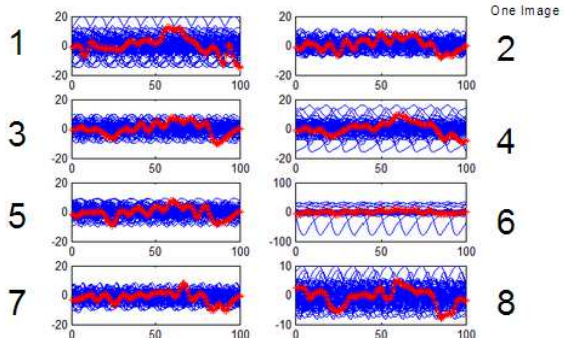


Fig 9. One second of EEG response for each channel. In each plot the 100 vectors associated to each windows of one image. In red, the target reaction, in blue the 99 distractors for one of the 32 repetitions.

The waveforms obtained are similar between the targets and the distractors. Whilst, there exists related work in the state of the art to analyze the single trials and extract the proper features by wavelet representations and complex methods, these are out of the scope of this thesis.

5) Averaged Trials

The average of the 32 feature vectors of each window was computed, following the same idea to find the ERPS of the flashes presented in Section IV.A.

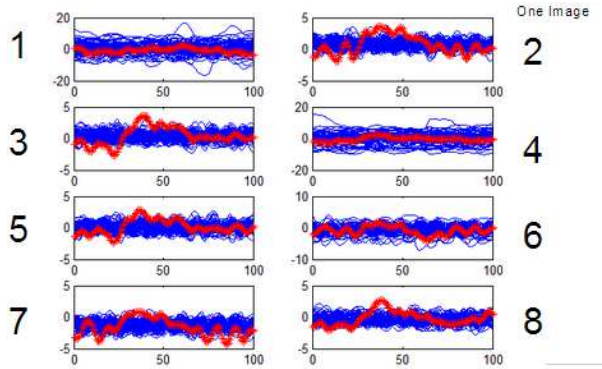


Fig 10. 1 second of averaged EEG response for each channel. In each plot the 100 averaged vectors associated to each windows of one image. In red, the averaged reaction for 32 examples of target window, in blue the 99 averaged reactions for the distractors.

Figure 10 indicates now that the peaks of averaged targets in channels 2, 3, 5, 7 and 8 may indeed be distinctive to the averaged distractors if considering the amplitudes. This observation suggests that the two patterns can be discriminated through machine learning techniques.

6) Feature Extraction

The absolute value of the signals and the mean of 96 targets and 96 distractors from a random window are plotted in Figure 11.

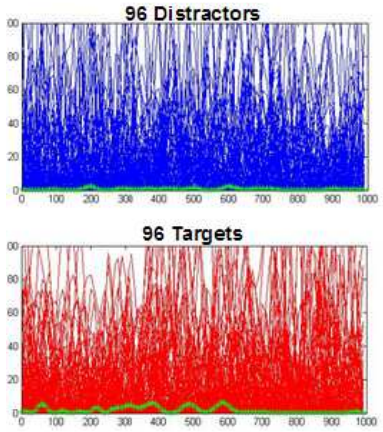


Fig 11. Absolute value for the EEG response associated to 96 distractor and 96 target of one channel. In green, the mean of the values.

It can be observed that the mean of the absolute values of the distractors and the targets is mainly different during the first 600ms. For this reason, it was decided to characterize each window directly with the energy value of the EEG response from the 0 to the 600ms after the visual stimuli.

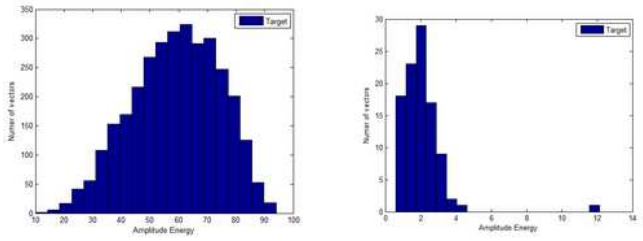


Fig 12. For one image (100 windows) and considering EEG channel 2: The histogram of the values of the energy computed. On the left, the values computed over the single trials; on the right the values computed over the average of the 32 trials.

Considering only one channel and computing the energy feature for all the single trials of one image, a similar value is obtained for all the windows (Fig. 12, left). Meanwhile, when the 32 trials of each window are averaged, a clear distinction between the distractors and the target window is obtained (Fig. 12, right). This result means that it is easier to distinguish the signals between averaged EEG responses of the windows presented than analyzing the single trials.

Furthermore, and focusing on only one channel, the fact of having a single value per window allows us to generate the first EEG map of the synthetic images based on the energy of the averaged signals. These initial EEG maps are show in Figure 12. The position of the target window (in white) is clearly distinguishable in 3 of the 4 images only by considering one of the channels of the EEG device and without any classification algorithm.

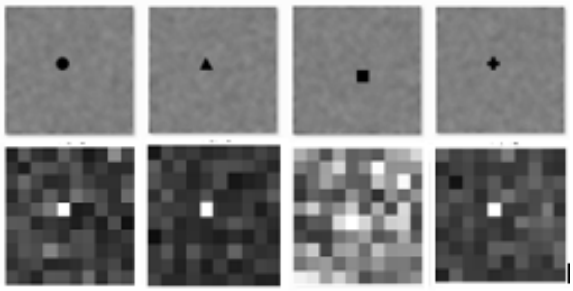


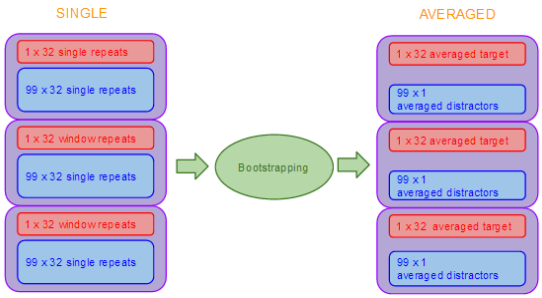
Fig 12. Synthetic images presented and their EEG map based on the value of the energy for the averaged trials.

In order to compute the maps, the value of the energy was normalized to 1 by dividing all the scores obtained of all the windows by the maximum energy value of the image. This post-processing is assuming that at least one window contains the object of interest. This normalization should not occur if the image may not contain any object.

7) Bootstrapping for the Generation of Averaged Data

Previous results indicate that averaging the signals is a good practice to identify the target windows. However, averaging reduces the amount of target examples from 32 single trials to just 1 averaged trial. This amount is not enough data to train the classifier algorithm.

Acquiring more user data was not feasible because of the limited access to the acquisition equipment and, even more importantly, the slow and stressing process of data acquisition for volunteers. For this reason, a bootstrap aggregation with no replacement was applied to generate 96 new examples of averaged EEG reactions. This technique generates a new sample by averaging 16 target examples randomly selected from the 32 available.



14. Boosting technique to generate averaged data.

The boosting is applied as well to generate averaged distractors for each of the corresponding 99 images.

8) Support Vector Machine (SVM)

The algorithm selected to classify windows was a SVM with a linear kernel from the package LibSVM for Matlab. This library provides the classification label for each instance and the probability value that the instance belongs to the predicted

class. The value of the probability is considered instead of the binary classification in order to generate a grey scale EEG map, similar to the saliency and segmentation maps used in related works.

The energy associated to each 0-600 ms window for each of the 8 channels is considered to define a feature vector of dimension 8. This method of channel fusion will let the SVM automatically learn which of the channels have the most relevant information and which ones are mainly noise to efficiently combine their outputs.

The SVM classifier was trained with 96 feature vectors of target windows and 96 from the distractors, corresponding to 3 images used for training. Each training image provides 32 target feature from its target window, but the 32 distractor features from a random sampling among the 99 distractor windows available.

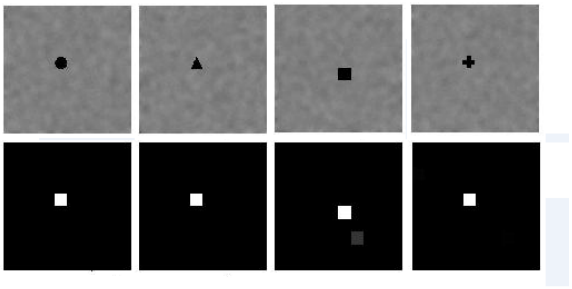


Fig 13. EEG maps generated by the probability value provided by the SVM.

The model obtained is tested with the 100 feature vectors of the remaining image. A cross validation approach is applied by running the experiments with the four possible combinations of 3 train + 1 test.

Assuming that the output has to have one object, the EEG map is computed by taking the probability value of the window classified as a target window, and normalizing this score by the maximum value obtained. The results obtained are shown in Figure 13, which clearly shows that it is possible to detect the object of interest.

VI. CONCLUSION

The work developed did not reach the original goals of the project: use the values obtained from EEG maps be compared to saliency maps for real images and used as the basis of a segmentation algorithm. This was because it became clear in this course of this project that these were extremely ambitious objectives that would require significantly more time. Nevertheless, the results obtained from the synthetic images provide evidence that BCI devices could in principle be used to locate an object into an image, this result represents a solid basis to perform more trials with real images in order to achieve the original objectives in the future.

The innovation of the work is the simplicity of the system: Simply by extracting the value of the energy of the averaged EEG waves, and combining the values obtained from each of the 8 channels via to train a SVM with a lineal kernel it is possible to locate the object in the synthetic images, whilst keeping the user free of any interaction during stimuli presentation.

The main weaknesses are that it only has been proved with a "simple" images, where the object was fitted in only one of the windows of the image instead of tried in real images, where having the object partially included in different windows becomes more challenging. This opens a huge range of variables like size of the object, size of the window, percentage of object displayed in the window, number of repetitions of the window, etc. that may affect in the issue of detecting the target windows.

The system works well when the averaged signals are considered. The direct consequence of this is that, even by using the boosting technique to generate more examples of data, the number of the image repetitions is high (in this study it is required 32 repetitions of each image to succeed in the classification of the windows).

Future work should study the extraction of better features that may reduce the number of image repetitions and focus in analyzing real images.

ACKNOWLEDGMENT

I would like to acknowledge to my supervisor Xavi Giro for his constant support and advices, to also my supersivors Kevin McGuinness and Noel O'Connor for their guidance throughout the entire project. To Ramya Hebbalaguppe for being always avaliable to help me and to Michael Keane for his guidance and good advices in the EEG world. Also to all the volunteers who left me their brain in the trials, and all the people who have been interested and have dedicated some time to this project. Thank you!

REFERENCES

[1] N. Ouerhan, R. v. Wartbur, H. Hügli and R. Müri, "Empirical Validation of the Saliency-based Model of Visual," in *Electronic Letters on Computer Vision and Image Analysis*, 2003, pp. 13-24.

[2] R. Hebbalaguppe, K. McGuinness, J. Kuklyte, G. Healy, N. O'Connor and A. F.Smeaton, "How interaction methods affect image segmentation: user experience in the task," in *The 1st IEEE Workshop on User-Centred Computer Vision (UCCV)*, Tampa, Florida, U.S.A., 2013.

[3] J. Wang, E. Pohlmeyer, B. Hanna, Y.-G. Jiang, P. Sajda and S.-F. Chang, "Brain state decoding for rapid image retrieval," in *MM '09 Proceedings of the 17th ACM international conference on Multimedia*, New York, 2009.

[4] Alan F. Smeaton, Graham Healy, "Optimising the number of channels in EEGAugmented image search," *BCS-HCI '11 Proceedings of the 25th BCS Conference on Human-Computer Interaction*, pp. 157-162, 2011.

[5] Nima Bigdely-Shamlo, Andrey Vankov, Rey R. Ramirez and Scott akeig, "Brain Activity-Based Image Classification from Rapid Serial Image resentation," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 16, pp. 432-441, octubre 2008.

[6] G. Healy, "An Analysis of EEG Signals Present During Target Search,", PhD thesis, Dublin City University, 2012.

[7] M. Toscani, T. Marzi, S. Righi, M. Viggiano and S. Baldassi, *"Alpha waves: a neural signatureof visual suppression"* . Experimental Brain research, Volume 207, Issue 3-4, pp 213-219, 2010.

[8] Suresh R. Devasahayam, "Signals and systems for bioengineers: A Matlab-Based introduction", Second Edition

[9] C. Bandt, M. Weymar, D. Samaga and AO Hamm, "A simple classification tool for single-trial analysis of ERP components", Psychophysiology, Vol. 46, No. 4 pp 747-757, 2009.