COMBINING OBJECT DETECTION AND BRAIN COMPUTER INTERFACING: THE MIND SCENE SPELLER

Arne Robben, Nikolay Chumerin, Nikolay V. Manyakov, Adrien Combaz, Marijn van Vliet and Marc M. Van Hulle

K.U.Leuven, Laboratorium voor Neuro- en Psychofysiologie, Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium

ABSTRACT

In this paper we propose an application which combines two research disciplines: object detection and brain-computer interfacing. It is in particular useful for patients suffering from a severe motor impairment which prevents them to interact with their surrounding environment. The application shows an image of e.g., the room of the patient, on a computer screen and searches for instances of certain objects in the image. When these are found, a flashing dot appears on top of them, flickering in a fixed but different frequency for each object. Meanwhile, brain-activity (EEG) is recorded. Selecting an object can then be achieved by looking at the corresponding flashing dot: the application processes the EEG-readings and identifies the frequency embedded in the signal (SSVEP decoding). Therefore it can conclude on the object the subject was looking at. In this way a patient can (re)gain interaction with his or her environment.

I. INTRODUCTION

In the last decade, research on *object detection* has become a flowering branch in the domain of computer vision. The task is, given an image, to conclude on the presence of a specified object and, if present, to determine its location in the image. Research on detectors for faces, cars, motorcycles, pedestrians, road signs and many more, already led to successful and reliable applications [1], [2], [3], [4].

Most techniques are based on the matching of local features of an image to a database of features (a codebook) derived from a training-set of an object class. Among these local features are grayscale patches, Haar-like features, local shape contexts, SIFT, SURF, etc... [2], [5], [6], [7].

For this application we made codebooks of SIFT-descriptors. These local features are 128-dimensional vectors, proposed by David Lowe in 1999 and are constructed in such way that they are invariant to scale and rotation (Scale Invariant Feature Transform abbreviates to SIFT). Besides this, they are also claimed to be partially invariant to a substantial change of affine distortion, change in 3D viewpoint, addition of noise and illumination [8]. We constructed 3 codebooks: one for a coffee-thermos, one for a cup with an apple-print, and one for a white CRT-monitor (see Fig. 1.), objects which might appear in everyday environments.

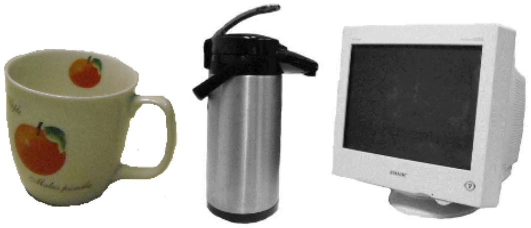

Fig. 1. The cup, thermos and monitor for which codebooks were constructed.

Our application connects the domain of object detection with the research topic of *Brain-computer interfaces* (BCIs), resulting in a combination which has, to the best of our knowledge, never been looked into before.

Over the last few years, research on BCIs is witnessing a tremendous development (see, for example [9]). These interfaces are able to directly read out brain activity and thus establish a communication pathway between the brain and a computer, bypassing the need for muscular activity. As such, BCIs can significantly improve the quality of life of patients suffering from impairments as amyotrophic lateral sclerosis, stroke (CVA), brain/spinal cord injury, multiple sclerosis, etc.

AR and AC are supported by a specialization grant from the Agentschap voor Innovatie door Wetenschap en Technologie (IWT, Flemish Agency for Innovation through Science and Technology).

NC is supported by the European Commission (IST-2007-217077).

NVM is supported by the Flemish Regional Ministry of Education (Belgium) (GOA 10/019).

MMVH is supported by research grants received from the Excellence Financing program (EF 2005) and the CREA Financing program (CREA/07/027) of the K.U.Leuven, the Belgian Fund for Scientific Research - Flanders (G.0588.09), the Interuniversity Attraction Poles Programme – Belgian Science Policy (IUAP P6/054), the Flemish Regional Ministry of Education (Belgium) (GOA 10/019), and the European Commission (STREP-2002-016276, IST- 2004-027017, and IST-2007-217077), and by the SWIFT prize of the King Baudouin Foundation of Belgium.

In principle, two kinds of BCIs can be discerned: invasive (intra-cranial) and non-invasive ones. The former are characterized by implanted micro-electrodes, mostly in the premotor-or motor frontal areas or into the parietal cortex (see, for example [10]), the latter mostly work with *electroencephalograms* (EEGs) recorded from the scalp.

There are many kinds of EEG-BCIs, categorized according to the paradigm that is behind the BCI. A first group relies on the $P300$ *event-related potential* (ERP) [11]. A brain potential is prominent in the parietal cortex as a response to infrequent but preferred stimuli in contrast to non-preferred high-frequent stimuli (also called the oddball paradigm). A second group is based on the detection of mental tasks (imagination of limb movement, subtraction, word association, etc...) which are discovered through *slow cortical potentials* (SCP) [12], *readiness potentials* [13] and *event-related desynchronization* (ERD) [14].

The third group relies on the detection of *steady-state visual evoked potential* (SSVEP) responses and is also used for our application. If a visual stimulus flickers at a sufficiently high rate ($\geq$ 6 Hz), individual transient visual responses overlap, resulting in a *steady state* signal, observable mostly in the occipital area. [15]. Not only the stimulus frequency f can be discovered in the signal, the harmonics $2f$ and $3f$ are also often embedded in the signal.

When there are multiple targets flickering in different frequencies it intuitively suffices to look at the maxima of the Fourier transform of the EEG-signal and decide with this information on which frequency the subject was focusing (as illustrated in Fig. 2.). It is however often not that easy. One problem is that the amplitude of a typical EEG in the spectral domain is inversely proportional to the frequency. The major problem though, is due to the nature of the EEG-recordings: a lot of noise and other on-going brain activity are present in the signal. Standard techniques for dealing with this problem record over a long time interval, average over several time intervals[16] or make use of preliminary training. In this application we use a different approach, inspired by the method proposed by [17], which does not require a preliminary training stage.

Two main results of our studies will be treated: the performance of our detection system, and the SSVEP-decoder. A study with X subjects was carried out to estimate the accuracy of the SSVEP classifier.

II. OBJECT DETECTION

II-A. Acquiring SIFT-features

Our object detector is mainly based on the matching of SIFT-features found in a scene-image to SIFT-features found from pictures from a cup, thermos and monitor under different viewpoints and stored in a 'codebook'.

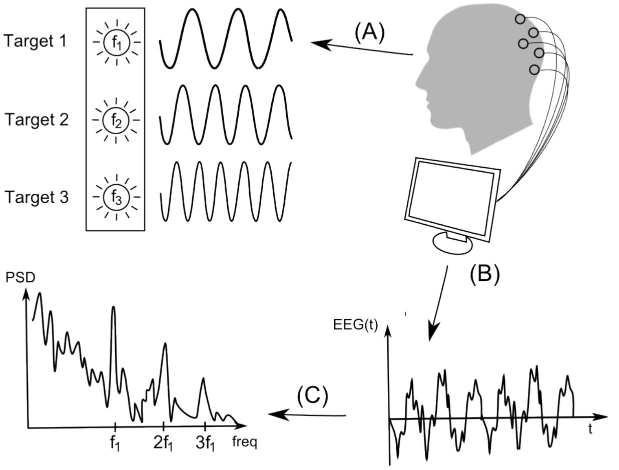

Fig. 2. Standard SSVEP-decoding approach: (A) a subject looks at Target 1, flickering with frequency f_1, (B) noisy EEG-signals are recorded, (C) taking the Fourier Transform over a sufficient large window shows peaks at $f_1, 2f_1$ and $3f_1$.

Acquiring SIFT-features from an image starts from detecting interest points (or keypoints) in this image, these are salient points in the image with rich local information. In order to achieve these interest points, a *scale space* is constructed by convolving Gaussians in variable scale with the input image and taking the difference of these convolutions for nearby scales (also called *Difference of Gaussians* (DoG)). The interest points correspond to local minima/maxima of the DoG images (for more details, see [8])

Next, based on local grayscale image properties, a dominant gradient direction is computed for each interest point in the scale where it was found. This is not only done for the interest point but also for every pixel in the neighborhood of this keypoint, giving more weight to nearby pixels then to pixels further from the interest point. Putting these orientations into local orientation histograms provides us with the entries which make up the 128-dimensional descriptor. By normalizing the descriptors to unit length, the effects of illumination change are said to be reduced (again, see [8]).

The images used in this study where all taken by a digital camera and have a resolution of 2304×3072 each. In large images many small gradients are detected even if these gradients are due to small texture variations, reflecting light, etc. What we really want are keypoints found on the level of the objects in the image: on the outline of objects, on logos on the object, on handles off the object, etc. This is where a first user-dependent parameter comes into play: how much downscaling is needed to get optimal performance, both for the scene-image as for the training-images of our objects.

We designed 3 detectors, one for each object. They only differ in the way the scene-image and training-images are scaled: the cup has for example smaller details then the thermos so we will downscale thermos-images more then cup-images. After some trial and error, the following downscaling was applied for thermos, cup and monitor images (both for training and scene-images): 225×300, 750×1000 and 527×700 respectively.

II-B. Training a detector

Around 15 training images where taken of each object, taken under different viewpoints. The objects where manually segmented from the background and downscaled like above. Each training-image in now filled with the object. It is however very likely that in scene-images this object will appear smaller then in the training images. Although the SIFT-descriptors are scale-invariant, we really want descriptors on the level of the object (like discussed above), therefore we set up a vector α *of reasonable scales* (in percentage) in which the object might occur in a scene image. For the monitor this is for example a linearly spaced vector between 0.30% and 0.52%, with 6 elements. For each scale α and for all training-images, SIFT-descriptors are computed and stored in a codebook $\mathbf{C}_\alpha$. Finally, the *center* $\mathbf{c}$ of each object in each training image is manually assigned and the location $\mathbf{loc_k}^\alpha$ of each keypoint $\mathbf{k}$ in the image on scale α, relative to $\mathbf{c}$, is stored (as illustrated in Fig. 3).

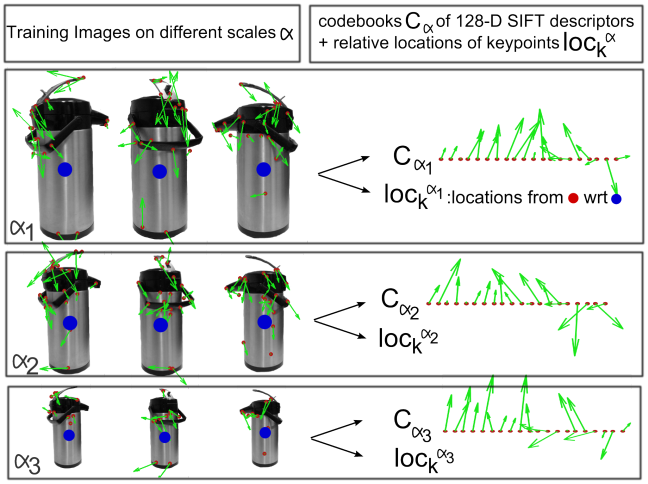
Training Images on different scales α
codebooks C_α of 128-D SIFT descriptors
+ relative locations of keypoints loc_k^α
C_{α_1}
$loc_k^{\alpha_1}$:locations from ● wrt ●
α_1
C_{α_2}
$loc_k^{\alpha_2}$
α_2
C_{α_3}
$loc_k^{\alpha_3}$
α_3

Fig. 3. Construction of codebooks and relative keypoint locations, from training images on different scales.

II-C. Detecting an object in a scene

Now the codebooks $\mathbf{C}_\alpha$ and the relative keypoint locations $\mathbf{loc_k}^\alpha$ are computed for an object, the recognition for this object in a scene-image can begin. First the scene-image is loaded in and rescaled as above. SIFT-descriptors are computed and matched for each α with the codebook $\mathbf{C}_\alpha$ (matching as in the sense of [8]). In order to recognize an object in the scene, a *voting space* $\mathbf{V}_\alpha$ is created for each scale α; this is a zero-valued matrix with the same size as the image. We will now fill these voting spaces with votes *for the center of the object* in the scene. The idea is to eventually compare all votes over all scales, find the maximum vote and compare it to a user-defined threshold. In this way we will be able to conclude on both the presence and the location of the object.

When a descriptor $\mathbf{k}'$ in the scene matches with a keypoint $\mathbf{k}''$ in $\mathbf{C}_\alpha$, a vote can be cast for the center of the object in $\mathbf{V}_\alpha$: we just distract from $\mathbf{k}'$ the relative position to the center that we stored for $\mathbf{k}''$. The vote goes thus to the location: $\mathbf{k}' - \mathbf{loc_{k''}}^\alpha$. To incorporate the uncertainty of this vote we store a Gaussian $N(\mathbf{loc_{k''}}^\alpha, \sigma)$ in $\mathbf{V}_\alpha$ with σ around 0.10 times the height of the scaled-down scene image.

Iteratively, all keypoints in the scene-image are processed in this way (see Fig 4) and votes in $\mathbf{V}_\alpha$ are accumulated. Finally the maximal vote over all scales α is our guess for the location of the object and our application draws a colored dot on the object. The value of this maximum can be treated as a confidence measure for this guess, especially when it is normalized with respect to $\mathbf{V}_\alpha$. In our case we divide each Gaussian vote by its maximum value and by the number of found keypoints. If one position would get all votes the maximum of the voting space on this position would then be equal to 1. By comparing the weighted maximum over $\mathbf{V}_\alpha$ with a user-defined threshold $\mathbf{t}$, the detector is thus able to decide on the presence of the object in the scene-image.

III. SSVEP DECODING

III-A. EEG Data acquisition and filtering

The EEG recordings were performed using a prototype of an ultra low-power 8-channels wireless EEG system, which consists of an amplifier coupled with a wireless transmitter and a USB stick receiver, developed by IMEC[1]. The data is transmitted with a sampling frequency of 1000 Hz for each channel. We used a brain-cap with large filling holes and sockets for active Ag/AgCl electrodes (ActiCap, Brain Products). The recordings were made with eight electrodes located on the occipital pole (covering the primary visual cortex, where the SSVEP is most prominent), more precisely at positions P3, Pz, P4, PO9, O1, Oz, O2, PO10, according to the international 10–20 system. The reference electrode and ground were respectively placed on the left and right mastoids.

The raw EEG signals are filtered above 3 Hz, with a fourth order zero-phase digital Butterworth filter, so as to

[1] Interuniversity Microelectronics Centre (IMEC), *http://www.imec.be*

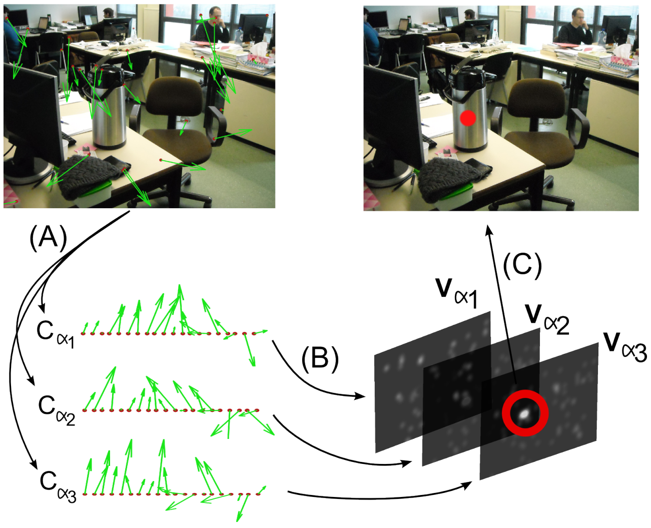

Fig. 4. Detecting the thermos: (A) SIFT-features are extracted and matched against the codebooks, (B) votes for the center of the thermos are casted, (C) the maximal vote is selected as our guess for the object location.

remove the DC component and the low frequency drifts. A notch filter is also applied to remove the 50 Hz powerline interference.

III-B. Experiment design

X healthy subjects (x male, x female, age x-x, x left handed, x right handed) participated in the experiment. One session lasted around x hours in order to maintain the concentration of the subjects. Because our application was trained on 3 objects, 3 flickering dots were presented to the subject, flickering in different frequencies f_1, f_2 and f_3. The visualization of the stimuli was implemented via the *Psychtoolbox*[2] for Matlab (see Fig 5). The subjects were asked to focus on each dot for x seconds, a pause was taken and the session was repeated x more times. During preliminary experiments, it became obvious that the choice of stimulation frequencies, in terms of accuracy, is very subject dependent. Therefore a calibration stage was first introduced showing a wide range of frequencies each for x seconds. By visually inspecting the spectrogram, 3 prominent frequencies could be chosen.

III-C. Spatial filtering and classification

We designed a spacial filter in the sense of [17]; a linear combination of the signals from our 8 channels is sought which decreases the level of noise in our frequencies of interest: the target frequencies f_1, f_2, f_3 and their harmonics $2f_1, 3f_1, 2f_2, 3f_2, 2f_3$ and $3f_3$ (also called *Minimum Energy Combination*). This can be done in 2 steps. First consider the $(T \times 8)$ matrix $\mathbf{X}$ containing the recorded EEG data of duration T for each channel in the columns. A $(T \times 18)$ matrix $\mathbf{A}$ is then constructed with the functions $sin(2\pi h f_i t)$ and $cos(2\pi h f_i t)$ in its columns, in the time-moments $t \in 1, ..., T$ and where $h \in \{1, 2, 3\}$ denotes the harmonics of $f_i, i \in \{1, 2, 3\}$. By multiplying $\mathbf{X}$ with the $T \times T$ projection matrix $P_A = A(A^T A)^{-1} A^T$ and subtracting this matrix from $\mathbf{X}$ gives us $\tilde{\mathbf{X}} = \mathbf{X} - P_A \mathbf{X}$, a matrix like $\mathbf{X}$ but without the information about the target frequencies and their harmonics. $\tilde{\mathbf{X}}$ can be considered as the components of the original signal which are not related to the visual stimulation.

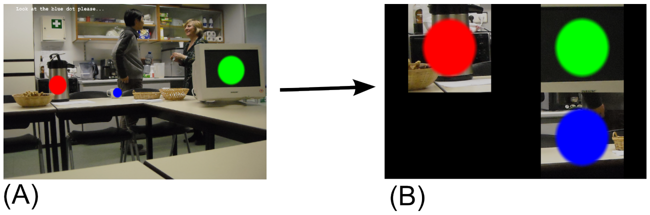

Fig. 5. After detecting the objects in the scene image (A), the dots are enlarged and positioned as in (B), now the actual stimulation can begin.

The second step is to find a linear combination of $\mathbf{X}$ which minimizes the variance of these non-interesting components. This can be obtained by performing a *principle component analysis* on $\tilde{\mathbf{X}}$, these principle components correspond to the eigenvectors of the covariance[3] matrix $\Sigma = E[\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}]$. The first principle component v_1 points in the direction of the maximum variance of the data, the second component v_2 lies in the direction of maximum variance in the space orthogonal to v_1, etc. In this way, an orthogonal projection of the data on the first principle components takes as much variance as possible to a lower dimensional space. The other way around is in this case of our interest: projecting on the *last* principle components represents the data in a lower dimension with a lot *less* variance. Because we compute these principle components for $\tilde{\mathbf{X}}^T$ (the matrix containing non-interesting information), the resulting projection $8 \times k$ matrix V_k (columns are k-last eigenvectors) times $\mathbf{X}$ corresponds to the linear combination of $\mathbf{X}$ which minimizes the variance of these non-interesting components, we write: $\mathbf{S} = \mathbf{X} V_k$. We choose k such that $\sum_{i=8-k}^{8} \lambda_i / \sum_{i=1}^{8} \lambda_i < 0.1$, where λ_i is the eigenvalue corresponding to the eigenvector $v_i, i = 1, ..., 8$.

To classify the stimulation frequency, test statistics T_i are calculated for each target frequency $f_i, i = \{1, 2, 3\}$

[2] *http://psychtoolbox.org*

[3] $E[\cdot]$ denotes the statistical expectancy. By definition $\Sigma = E[\tilde{\mathbf{X}}^T \tilde{\mathbf{X}}] - E[\tilde{\mathbf{X}}^T] E[\tilde{\mathbf{X}}]$, but because of our filtering method the data is centered around zero, the last term drops.

and by making use of $\mathbf{S}$. Formally

$$T_i = \sum_{h=1}^{3} \sum_{s=1}^{k} \frac{\mathbf{P}_{h,s}(f_i)}{\sigma^2_{h,s}(f_i)},$$

where $\mathbf{P}_{h,s}(f_i)$ and $\sigma^2_{h,s}(f_i)$ are estimates of respectively the power and the noise of target frequency f_i in its harmonics (index over h) and estimated by means of the s-th column of $\mathbf{S}$. Classification is then as simple as taking the maximum of these test statistics $T_i, i = \{1, 2, 3\}$.

$\mathbf{P}_{h,s}(f_i)$ can be computed as

$$\mathbf{P}_{h,s}(f_i) = \sum_{t=1}^{T} (s(t) sin(2\pi h f_i t) + s(t) cos(2\pi h f_i t)),$$

with $s(t)$ the s-th column of $\mathbf{S}$ (which is simply the discrete Fourier transform magnitude at frequency hf_i).

Estimating the noise power is a harder thing to do. Following [17] we used an autoregressive model on $\mathbf{S}$ because it can be considered as a filter (working through convolution), in terms of ordinary products between transforms of signals and filter coefficients in the frequency domain. Since we assume that the prediction error in autocorrelation model is uncorrelated white noise, we have a flat power spectral density for it with magnitude as a function of the variance of this noise signal. Thus, the Fourier transformations of the regression coefficients a_j (estimated, for example, with use of Yule-Walker equations) show us the influence of the frequency content of particular signals into the white noise variance. More formally, we have

$$\sigma^2_{h,s}(f_i) = \frac{\pi T}{4} \frac{\tilde{\sigma}^2}{|1 - \sum_{j=1}^{p} a_j \exp(-2\pi\sqrt{-1} j h f_i / F_s)|},$$

where T is the length of the signal, $\tilde{\sigma}$ is an estimate of the variance of white noise, p is an order of regression model, F_s is the sampling frequency (1000 Hz).

IV. RESULTS AND DISCUSSION

IV-A. The object detector

Measuring the performance of a detection system is not trivial. Depending on the lightning conditions of a scene image, the amount of detail of the object, the amount of detail in the scene image responsible for triggering false interesting points, the scale of the sought object in the scene, and many more factors, the object detector will perform better or worse. Comparison with other studies is also not straightforward; most widespread datasets contain images of faces, cars, pedestrians, etc. The goal of our application is though to detect objects which could be found in the near environment of a patient. Therefore we took for the cup, monitor and the thermos respectively $100, 97$ and 120 scene images, taken inside our lab under both natural and artificial light, in different rooms and viewpoints and always with other objects in the neighborhood. The scale of the object inside the scene image was a little restricted, for example: the height of the monitor was always around 0.3 to 0.5 times the height of the image. By doing so, the vector α of reasonable scales in section II-B could be constructed.

By visually inspecting the output of the object detectors we discerned the correct detections (true positives) as the detections where the winning vote for the object center (as defined in section II-C) is a pixel belonging to the object. If the vote does not belong to the object it is a false positive. Like in [2] the number of true and false positives detections are used to compute the precision and recall and the receiver operator characteristic (ROC), while varying the threshold for detection as proposed in section II-C. The result is shown in figure 6

Again it should be well emphasized that this measure for the performance is not absolute. It remains a big challenge to constrain the parameters which influence the accuracy of an object detector and at the same time be general enough to find the object in a large amount of different scene images. In testing the classifier we found that the choice of α (the vector of "of reasonable scales" from section II-B) is extremely influential. Despite these difficulties, our results do show the potential of these kind of object detectors for our application. One nice addition would be to implement a generic training phase where new objects can be trained in order to increase the environmental interaction of a specific patient.

IV-B. The SSVEP classifier

During a calibration stage 3 frequencies where chosen for each subject and data was recorded for x seconds (see II). The accuracy of the SSVEP classifier was then off-line assessed in terms of the duration of the EEG-recordings. The accuracy can be described by the number of correct predictions divided by all predictions made by the classifier. These results are brought together in figure 6.

When there would be more object detectors available, the selection task by means of different frequencies becomes more difficult: the minimal energy combination method for 4 or more targets is less precise then for 3 targets and it is harder to find 4 or more different distinct frequencies for each subject. One solution for this can be to group objects together and do a tree search; if one group of objects is selected a new selection phase begins where each individual object from this group becomes selectable. Another strategy to handle more targets is not only to incorporate the magnitude of a selection of frequencies but also its phase. For example: in [18], 8 targets where shown, flickering in the same frequency but each with a different phase-shift. Either of the methods or a combination of both

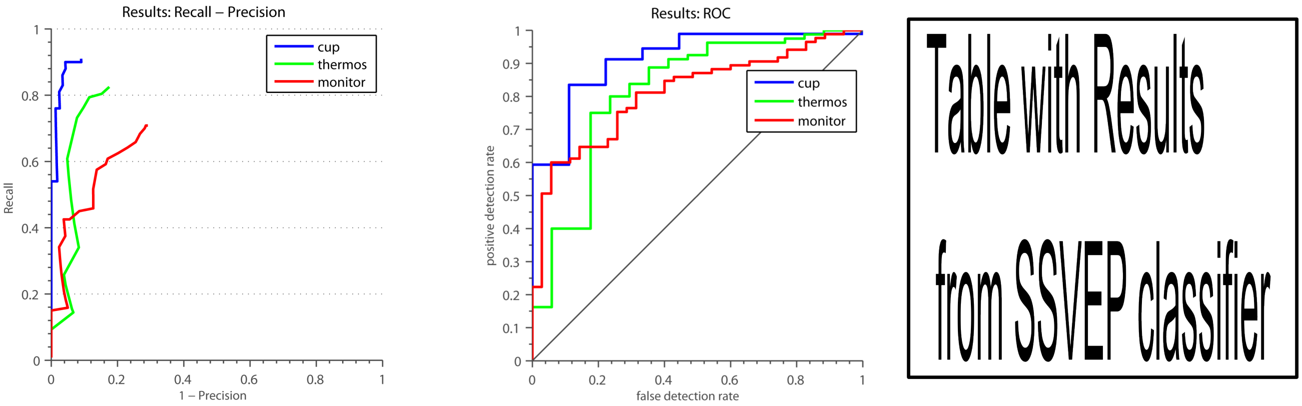

Fig. 6. Left and middle figure: accuracy of the object detectors when varying the threshold as described in section II. Right figure: accuracy of the SSVEP classifier as a function of the data duration t in seconds.

can therefore handle an increase in the number of object detectors.

V. CONCLUSION

A new application was presented which can improve the interaction of motor impaired patients with their environment. It allows the patient to select objects by means of object detectors and brain computer interfacing. Improvements in both research domains will directly lead to an improvement of our application.

VI. ACKNOWLEDGMENT

The authors also grateful to Refet Firat Yazicioglu, Tom Torfs and Cris Van Hoof from the Interuniversity Microelectronics Centre (IMEC) in Leuven for providing with the wireless EEG system.

VII. REFERENCES

[1] P. Viola and M. Jones, “Rapid Object Detection Using a Boosted Cascade of Simple Features,” *CVPR*, 2001.

[2] S. Agarwal and D. Roth, “Learning a Sparse Representation for Object Detection,” in *Proc. Seventh European Conf. Computer Vision*, 2002, vol. 4, pp. 113–130.

[3] B. Leibe, A. Leonardis, and B. Schiele, “Robust Object Detection with Interleaved Categorization and Segmentation,” *IJCV*, vol. 77, 2008.

[4] G. Piccioli et al., “Robust method for road sign detection and recognition,” *Image and Vision Computing*, vol. 14, pp. 209–223, 1996.

[5] S. Belongie, J. Malik, and J. Puzicha, “Shape Matching and Object Recognition Using Shape Contexts,” in *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2002, vol. 2, pp. 509–522.

[6] D. Lowe, “Object Recognition from Local Scale-Invariant Features,” in *Proceedings of the International Conference on Computer Vision*, 1999, pp. 1150–1157.

[7] H. Bay, A. Ess, T. Tuytelaars, and L. J. V. Gool, “Speeded-up robust features (SURF),” *CVIU*, vol. 110, no. 3, pp. 346–359, 2008.

[8] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] P. Sajda, K. Muller, and K. Shenoy, “Brain-Computer Interfaces [from the guest editors],” in *IEEE Signal Processing Magazine*, 2008, vol. 25, pp. 16–17.

[10] G. Santhanam, S. I. Ryu, B. M. Yu, A. Afshar, and K. V. Shenoy, “A high-performance brain-computer interface,” *Nature*, vol. 442, pp. 193–198, 2006.

[11] W. S. Pritchard, “Psychophysiology of P300,” *Psychological bulletin*, vol. 89, pp. 506–40, 1981.

[12] N. Birbaumer et al., “The thought translation device (TTD) for completely paralyzed patients,” in *IEEE transactions on rehabilitation engineering*, 2000, vol. 8, pp. 190–3.

[13] B. Blankertz, G. Dornhege, M. Krauledat, K. Mller, and G. Curio, “The non-invasive Berlin Brain-Computer Interface: fast acquisition of effective performance in untrained subjects,” *NeuroImage*, vol. 37, pp. 539–50, 2007.

[14] J.R. Wolpaw, D.J. McFarland, and T.M. Vaughan, “Brain-computer interface research at the Wadsworth Center,” in *IEEE Transactions on Rehabilitation Engineering*, 2000, vol. 8, pp. 222–226.

[15] G. Bin, X. Gao, Y. Wang, B. Hong, and S. Gao, “VEP-Based Brain-Computer Interfaces: Time, Frequency, and Code Modulations,” *Computational Intelligence Magazine, IEEE*, vol. 4, no. 4, pp. 22–26, 2009.

[16] N. Manyakov, N. Chumerin, A. Combaz, A. Robben, and M. Van Hulle, “Decoding SSVEP responses using time domain classification,” in *Proceedings of the 2nd International Conference on Neural Computation*, 2010, pp. 376–380.

[17] O. Friman, I. Volosyak, and A. Gräser, “Multiple Channel Detection of Steady-State Visual Evoked Potentials for Brain-Computer Interfaces,” in *IEEE Transactions on Biomedical Engineering*, 2007, vol. 54, pp. 742–750.

[18] et al. P. L. Lee, “An SSVEP-Actuated Brain Computer Interface Using Phase-Tagged Flickering Sequences: A Cursor System.,” *Annals of Biomedical Engineering*, vol. 38, no. 7, pp. 2383–2397, 2010.