Vision: Smart Home Control with Head-Mounted Sensors for Vision and Brain Activity

Pieter Simoens
Dept. of Ind. Tech. & Constr.
Ghent University
V. Vaerwyckweg 1
B-9000 Gent, Belgium
psimoens@intec.ugent.be

Elias De Coninck
INTEC - iMinds
Ghent University
G. Crommenlaan 8 bus 201
B-9050 Gent, Belgium
edconinc@intec.ugent.be

Thomas Vervust
ELIS - CMST
Ghent University
Technologiepark 914
B-9052 Zwijnaarde, Belgium
thomas.vervust@elis.ugent.be

Jan-Frederik Van Wijmeersch
Ghent University
G. Crommenlaan 8 bus 201
B-9050 Gent, Belgium

Tom Ingelbinck
Ghent University
G. Crommenlaan 8 bus 201
B-9050 Gent, Belgium

Tim Verbelen
INTEC - iMinds
Ghent University
G. Crommenlaan 8 bus 201
B-9050 Gent, Belgium

ABSTRACT

Today, an increasing number of household appliances is being connected to the Internet to form a smart home. Intelligent control algorithms in the cloud adapt the configuration of this Internet-of-Things to our daily routines and personal preferences. Frequently, there are unforeseen situations where the control algorithms will not capture the actual desired configuration. In these cases, the user must intervene in the control algorithms and manually adjust the connected object's setting. Browsing to the appropriate web service or launching the vendor-specific companion app for even a simple interaction like lowering the temperature setting is a tedious process.

In this paper, we report on our early insights in building a mobile system that provides a common, intuitive interface to all actuators in the smart home. Using a head-mounted camera and a commercial Emotiv EEG neuro-headset, we let the user configure the IoT by merely looking at an object and performing a related facial expression. This way, users only need to look at an object and think about the desired action. We leverage on the home cloudlet for the compute-intensive signal processing for object detection.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems; I4.9 [**Image Processing and Computer Vision**]: Applications

Keywords

Emotiv EEG, Google Glass, wearable, Internet-of-Things, cloudlet

MCS'14, June 16, 2014, Bretton Woods, New Hampshire, USA.

http://dx.doi.org/10.1145/2609908.2609945.

1. INTRODUCTION

An increasing number of common household devices is being connected to the Internet. The set-top box was one of the first to be equipped with an Internet connection to enable interactivity, but recently also smaller devices become connected. There are many examples of already commercially available devices, like thermostats [12], light switches [2], LED lights [14], etc. Together, this Internet-of-Things (IoT) results in a smart home environment where the configuration of IoT actuators (e.g. desired temperature or light level) is adjusted to the preferences and routines of its residents.

Automated control algorithms are typically implemented as rule-based systems that perform well under expected situations. Real-time context construction has however largely remained unaddressed by the IoT research community [13]. Humans cannot be completely casted in rules, and the home resident's desires might change instantly. In these unanticipated situations the user will want to manually override the automated object setting, at least temporarily.

The configuration and settings of connected objects are typically managed in one of two ways: by web services in the cloud, or by a companion app on the mobile device of the resident. The commercial reality of a multi-vendor environment results in a modern version of the 'basket of remotes' problem [7]. For each device, users need to remember the correct URL or launch a particular vendor's app. Even a simple action like adjusting the room temperature thus requires a lengthy and awkward operation to perform. This commercial reality is far from Mark Weiser's vision of users interacting intuitively with technology disappeared in the environment [20].

Clearly, smart homes must be configurable via an intuitive interface, common to all actuators, that enables ad-hoc interactions in an unobtrusive way. We are currently building a prototype that meets these requirements. We combine the video feed from the camera integrated in smart glasses like Google Glass [8] with brain activity measurements captured with the Emotiv EEG neuro-headset [5]. The idea is that users can adjust an actuator by looking at the object at hand and performing a cognitive action to perform the actual operation. Users could then simply increase the room

temperature by looking at the thermostat and cognitively 'lifting' the object.

Figure 1 illustrates our vision in more detail. From the video feed, we detect the objects that are currently in view. Simultaneously, the captured cognitive input is analyzed and translated into the appropriate commands to manipulate the actuator in view. Both the smart glasses and the EEG headset lack the required computational resources and battery autonomy to perform continuous real-time signal processing. Rather, the heavy computation is offloaded to the user's smartphone or to the home cloudlet [15].

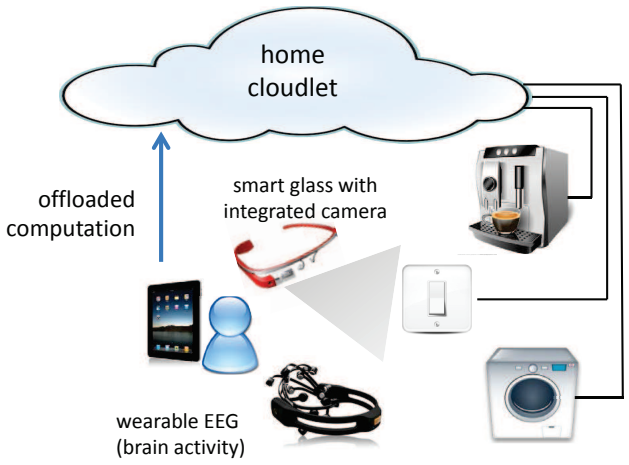

Figure 1: Users control the IP-connected objects in view via cognitive control. Heavy computation is offloaded to the home cloudlet.

Our prototype aligns to recent research directions that aim to provide a more natural human-computer interaction. Brain-computer interfaces have been mainly investigated in the context of disabled persons, e.g. to control a wheelchair [10] or speller application [18].

Our early prototype is composed of two head-mounted devices: a pair of smart glasses and the commercially available EEG neuro-headset, combined with a software implementation using existing frameworks like the Emotiv SDK and OpenCV for visual processing. The goal of this paper is to share our insights in the design and implementation choices of our prototype.

2. CAPTURING THE USER'S INTENT

A user's gaze represents the objects or persons that have his full attention. Like it is natural to look at the person you are talking with, it is normal to look at an object while manipulating it. We design our IoT interface on this premise. In a fraction of a second, the user must be able to tell the system *which* object he wants to control, and *how* he wants the configuration to be updated. To detect both aspects, we use a continuous, simultaneous capture and analysis of first-person video and brain activity respectively.

2.1 User gaze

Users will find it very intuitive to look at an object they associate with the action they want their smart home to carry out. If they want the light to be turned on, they should look at the light switch, and to increase the heat they should direct their attention to the thermostat.

To continuously capture the user's gaze, we exploit the front-facing camera that is mounted on wearable devices like Google Glass. This research line of 'continuous mobile vision', as first articulated by Bahl [1], is motivated by breakthroughs in hardware miniaturization, battery management and computer vision algorithms.

The captured video stream is analyzed in real-time to detect relevant objects in each frame. The captured frames are matched with key images that we expect to be delivered by the IoT device manufacturer. In fact, the commercial photographs used in catalogs or on web shops could be well suited to serve as key images, since these typically show the object isolated on a neutral background.

Identification of an object in a single frame is no solid ground to decide that the user wants to interact with that object. Many objects can be simultaneously in view, or objects could be captured coincidentally in the background. Although the presence of the object is correctly detected, it is still a false positive for our system if the user does not necessarily want to interact with it. Via user studies, we will have to analyze additional decision heuristics. One example could be that the object should be in the center of the captured frames for a sufficient number of frames.

We can further increase the usability of the system by associating multiple objects to the same action. For example, users should be able to increase the heating by looking at the radiator as well as the thermostat, although it is only the thermostat that is IP-connected.

2.2 Cognitive state

Brain activity and facial musculature movements result in electric potential variations that can be captured by a number of non-invasive sensors on the head. The use of these devices to control mobile devices has already been demonstrated by Samsung [9]. Campbell et al. have developed Neurophone [3], using neural signals to control the address book dialing app on a mobile phone.

The SDK provided with the Emotiv headset allows to detect various facial expressions, affective state (frustration, excitement) and cognitive activities like 'push', 'pull', 'rotate' and 'lift'. At first sight, the latter category would provide the most intuitive user interface. For example, the user could then cognitively 'rotate' the thermostat he is looking at to adjust the temperature, or 'lift' the light switch. However, the detection of these cognitive events is based on weak ElectroEncephaloGraphy signals (EEG) signals. This makes their measurement much more difficult, especially on low-cost commercial devices with less performant hardware [18]. Moreover, the brain patterns for these actions are highly individual. Training the Emotiv is not always easy, which could be an effort that many users might not be willing to make.

The electrical activity produced by skeletal muscles (EMG) is much more robust to detect and requires no training. Therefore, in our current prototype we have only taken into account user input via facial expressions. As the performance of the Emotiv headset increases, we can rapidly extend our system with cognitive events without having to change the system architecture.

Enhancing the user acceptance to wear BCI headsets requires a sleek design and a comfortable application process.

Current headsets like the Emotiv EPOC EEG require the user to hydrate the sensors with a saline solution before usage. The large number of visible electrodes does give the device a medical connotation which users might not find acceptable for daily usage. Future releases of EEG headsets, planned later this year, will address both requirements. Dry electrodes are hidden in a modern cover, and claimed to match the accuracy of wet sensors [4].

2.3 Visual-neural correlation

Individually, both signals are too ambiguous to accurately capture the targeted object and the desired change in configuration. Our system correlates in real-time the feature descriptors extracted from both the video feed and the electronic brain signals.

When an object is detected in the first-person video, this only indicates that the object is within the user's gaze. A recognized object does not necessarily mean that the user actually wants to manipulate that object. For example, users may be talking to another person in the living room. When the smart glass camera recognizes the light switch in the background, this information does not suffice for the system to decide that the user wants the lights to be switched off.

Similarly, the neuroheadset will continuously detect brain activity that not always reflects interaction with an IoT object. These headsets only capture electric potentials originating from brain activity or facial musculature, and do not 'read the user's mind'. Translating these signal patterns to brain-computer interface methods requires correlating them with the current user's context. For example, applying a visual or auditory stimulus will result in a well-known universal brainwave pattern. Existing BCI approaches exploit this by sequentially highlighting possible actions (e.g. letters). If the pattern is observed, this must be correlated to the action that was highlighted. While the signal itself only indicates the presence of a visual stimulus, the *meaning* of that signal is highly context-specific. In our prototype, we capture this context by first-person video.

3. PROTOTYPE IMPLEMENTATION

The main building blocks of our current prototype system are depicted in Figure 2. The hardware set-up of our early prototype currently comprises an off-the-shelf webcam and an Emotiv EEG headset. As the current Emotiv SDK is not supported on mainstream mobile platforms, we use a conventional laptop running Ubuntu. The webcam is connected via USB to the laptop, and the Emotiv headset pairs via Bluetooth with a USB dongle. We implemented in C++, using existing libraries like Boost and OpenCV. The device-cloudlet communication is implemented with Apache Thrift. This mock-up allowed us to quickly set-up a functionally complete system for experimenting with the parameters of the system. In the next version, we plan to integrate a Vuzix smart glass and an actual high-end smartphone in the set-up.

The laptop runs a controller component that filters both input signals before they are streamed to the cloud. In particular, the fidelity of both signals is adjusted via the frame rate or the reporting frequency of cognitive events. The controller parameters are set in coordination with the components executing on the cloudlet. For example, when no relevant brain activity is detected, it could decide not to stream any video frame to the cloudlet, since apparently the user does not try to interact with any object. Also, we discovered that the Emotiv fires many identical events shortly after each other. Filtering redundant frames and events reduces the number of traffic uploaded to the cloud, which in turn leads to battery savings for the mobile device.

On the cloudlet, object recognition algorithms compare the received frames with the key images for their product. We must take into account scale variations between the reference images of the object and the frame, as the distance between a user and the object might vary. Currently, we do not take into account rotation invariance, since we assume only limited variations in the position of the smart glass because users will mostly keep their head in the same position.

If an object is detected, the cognitive instructions are mapped to actual device instructions. The device drivers may run on the cloudlet, or may simply be a proxy to the remote driver, e.g. via a REST API. The object detection module also provides feedback to the controller on the client. This feedback can be used to configure the signal filters (e.g. do not send EEG/EMG input as long as no object is detected). In future versions, we can use this feedback to make the controller launch a visual notification signal on the smart glass.

4. SYSTEM DESIGN CONSIDERATIONS

Our main motivation is to provide an intuitive control of connected IoT objects. This requirement percolates through all system design aspects. Using our system should have a minimal impact on the daily routine. With Google Glass having set the reference for modern looking smart glass, the main obstacle we see from a hardware perspective is the requirement to hydrate the Emotiv EEG sensors with a saline solution. As mentioned earlier, this will be solved using dry electrodes in future releases. Hence, we focus on the software aspects.

4.1 Unobtrusive BCI

Unobtrusiveness means that the interface should be intuitive and fast, and that the system can be used out-of-the-box.

A BCI system can send commands, controlled by brain activity and distinguished by EEG signal processing. BCIs are categorized based on the EEG brain activity patterns into three different types: event-related potentials, steady-state visual evoke potentials or the P300 component of event related potentials [16]. This last approach requires only a few minutes of training time, and can carry the highest information transfer rate. Event related potentials (ERPs) are the measurement of brain responses to specific cognitive, sensory or motor events. For example, the P300 is a positive peak in the EEG signal that is observed 300 ms after the onset of a stimulus that is unexpected or rare, which was also used in the Neurophone system [3]. Visual evoked potentials are elicited by sudden visual stimuli and the repetitive visual stimuli would lead to stable voltage oscillations pattern in EEG that are referred to as steady-state visual evoke potentials.

Cognitive actions like push, pull and rotate almost directly mimick the physical action that users must carry out. Turning on a light switch would require the user to cognitively lift the switch. To detect cognitive actions, the current version of the Emotiv EEG headset uses advanced classifiers

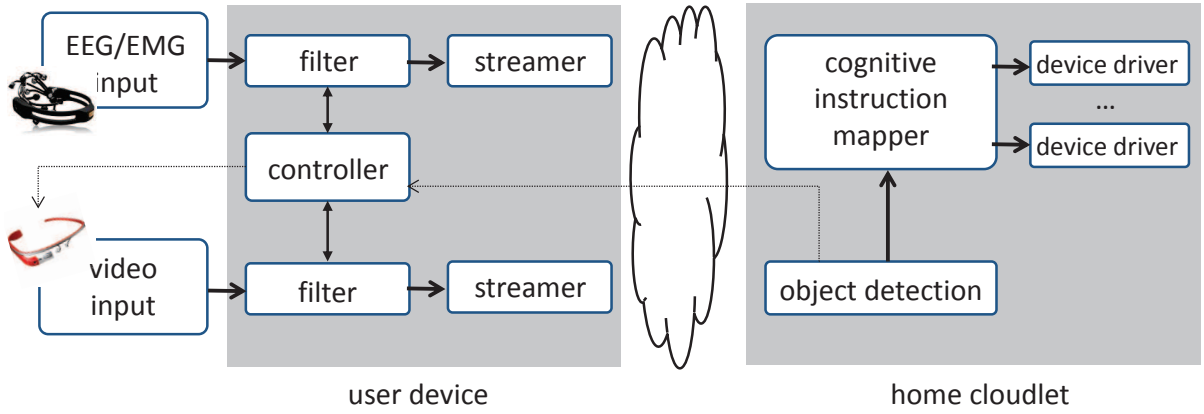

Figure 2: Block diagram of our current prototype. Neurological and visual signals are filtered and streamed to the cloudlet. On the cloudlet, the neurological signals are translated to object-specific instructions. If needed, feedback from the cloudlet is sent back to the controller.

and pattern recognition techniques to 'read' the user's mind. Unfortunately, this requires some training, as the brain signals associated with a particular cognitive action are highly individual [6]. Using these signals would force the user to manually train his system. From our experience, this is not a trivial task that would hinder the adoption of new systems.

For this reason, we decided to resort to the facial expression suite, like winking or frowning. Arguably, there is no intuitive correlation between the cognitive action and the physical manipulation. To mitigate this disparity, we can show the mapping of facial expressions to object manipulations on the display of the smart glasses. As soon as the object is detected, its interface is shown on the screen. Additionally, the overlay display can provide immediate feedback to the user on detected objects and registered cognitive actions. If desired, this user interface can be personalized; much like you can reprogram your remote control.

4.2 Energy

The energy cost of continuously streaming raw video and EEG signals over wireless and running classifiers or object recognition is challenging. Realizing a sufficiently long autonomy requires a combination of sensor hardware improvements and a rational system design that exploits available energy proportionality by reducing the fidelity of the captured signals as much as possible.

On the other hand, the sensing of brain activity consumes far less energy. The Emotive neuroheadset contains a built-in battery which is claimed to run for approximately 12 hours when fully charged. Reducing the fidelity of the brain events transmitted to the cloudlet is much more restricted. At most, we can only remove identical events that are detected in a very short time frame. Given the low data rate of this signal, the expected energy gains are far less than for the camera feed.

As of today, the developer APIs of mobile systems only provide little access to configure mobile cameras. At most, we can configure frame rate and resolution. The filter components in Figure 2 throw away redundant information (e.g. identic brain events) or unusable data (blurry frames). Moreover, the filter parameters are dynamically adjusted. The frame rate forwarded to the cloudlet can be lowered when an object is identified. The lower frame rate is sufficient to keep track if the user is still interacting with the same object.

Likamwa et al. [11] have studied the energy proportionality of image sensing by mobile cameras. They concluded that the overall sensing power indeed decreases with decreasing frame rate and resolution, but that the energy per pixel actually increases. As system capabilities and developer APIs mature, the authors expect the mobile image sensing to become more energy efficient.

Apart from the individual system optimizations for both sensors, we expect significant energy reduction by intelligently steering the each filter based on analysis of the other signal. In particular, an event detected in one stream can be used as trigger to restart the capture and transmission of the other stream. An important research question is the cognitive-visual causality to determine which signal can be used as trigger: i.e. will cognitive action be detected *before* or *after* the user has directed his gaze to the desired actuator?

4.3 Two-tier offloading

The captured neural and video signals must be heavily processed by classifiers. Object recognition is known to be a compute intensive task. Compared to research-grade EEG headsets, the signals of cheaper commercial devices are much noisier and require more sophisticated signal processing and machine learning techniques. In general, the real-time processing requirements are far beyond the hardware capabilities of smart glasses and neuro-headsets.

A common approach is to opportunistically offload heavy computation to nearby infrastructure [19]. In a home environment, there is ample infrastructure available that can assume the role of cloudlets: the settop-box, the desktop PC or the home automation server, which might already host the drivers for each of the connected objects in the smart home.

From the perspective of the wearable systems, we are provided with a two-tier cloudlet, consisting of the user's mobile device and the home cloudlet. The Emotiv EEG as well as the smart glasses are typically paired via Bluetooth to a nearby mobile device running a companion app, mak-

ing these devices a potential candidate as a first stage for offloading.

In our early prototype, we run the Emotiv processing on the mobile device, and offloaded all processing of the visual signal to the home cloudlet. Through experiments, we will need to evaluate the optimal trade-off in terms of energy consumption and performance between the local processing and offloading to the cloudlet. In general, the complexity of object recognition algorithms increases with the size of the set of candidate objects. In a smart home environment, the number of possible objects is limited and known in advance, so real-time object detection might be feasible on smartphones. We could even envision to reconfigure the object detection algorithm on the user device when the user enters a new room, which could be easily detected by the head-mounted camera [17].

5. CONCLUSIONS

In the near future, many actuators will be connected to the Internet and automatically controlled according to our preferences and daily routines. However, in many situations the user will want to manually override these systems. We propose a combination of head-mounted sensors to capture brain activity and visual information to realize an intuitive Brain-Computer Interface with the Internet-of-Things.

Our early prototyping work already revealed many design considerations. Many research questions arise from the combination of brain activity with first-person video, which we hope to address by setting up user pilot trials. In particular, we want to optimize our system design for performance and energy consumption. An important design tradeoff is whether the continuous object detection should be carried out on the smartphone, or on the home cloudlet. Another open question is how we can cope with multiple objects simultaneously in view. Lastly, we plan to investigate further the use of the glasses' overlay display.

6. ACKNOWLEDGEMENTS

Tim Verbelen is funded by a grant of the Fund for Scientic Research in Flanders (FWO-Vlaanderen).

7. ADDITIONAL AUTHORS

Additional authors: Maaike Op de Beeck (CMST-IMEC) and Bart Dhoedt (IBCN-iMinds)

8. REFERENCES

[1] P. Bahl, M. Philipose, and L. Zhong. Vision: Cloud-powered sight for all: Showing the cloud what you see. In *Proceedings of the Third ACM Workshop on Mobile Cloud Computing and Services*, MCS '12, pages 53–60, New York, NY, USA, 2012. ACM.

[2] Belkin. Wemo light switch, 2014. `http://www.belkin.com/us/p/P-F7C030/` [last visited on 2014-03-21].

[3] A. Campbell, T. Choudhury, S. Hu, H. Lu, M. K. Mukerjee, M. Rabbi, and R. D. Raizada. Neurophone: Brain-mobile phone interface using a wireless eeg headset. In *Proceedings of the Second ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds*, MobiHeld '10, pages 3–8, New York, NY, USA, 2010. ACM.

[4] J. D. Slater, G. P. Kalamangalam, and O. Hope. Quality assessment of electroencephalography obtained from a "dry electrode" system. *Journal of Neuroscience Methods*, 208(2):134 – 137, 2012.

[5] eMotiv. EEG System. `emotiv.com` [last visited on 2014-03-21].

[6] Emotiv. What kind of potentials does cognitive suite use?, 2014. `https://www.emotiv.com/ideas/forum/forum4/topic3876/` [last visited on 2014-03-21].

[7] J.-L. Gassee. Internet of Things: The "Basket of Remotes" Problem. `www.mondaynote.com/2014/01/12/internet-of-things-the-basket-of-remotes-problem/` [last visited on 2014-03-21].

[8] Google Inc. Glass. `www.google.com/glass/start/` [last visited on 2014-03-21].

[9] J. Hsu. Samsung imagines a future with mind-controlled tablets. *IEEE Spectrum*, 2013.

[10] T. Kaufmann, A. Herweg, and A. Kübler. Toward brain-computer interface based wheelchair control utilizing tactually-evoked event-related potentials. *Journal of neuroengineering and rehabilitation*, 11(1):7, 2014.

[11] R. LiKamWa, B. Priyantha, M. Philipose, L. Zhong, and P. Bahl. Energy characterization and optimization of image sensing toward continuous mobile vision. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '13, 2013.

[12] NEST. Nest home automation, 2014. `http://nest.com` [last visited on 2014-03-21].

[13] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos. Context aware computing for the internet of things: A survey. *Communications Surveys Tutorials, IEEE*, 16(1):414–454, First 2014.

[14] Philips. Hue personal wireless lighting, 2014. `http://meethue.com` [last visited on 2014-03-21].

[15] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies. The case for vm-based cloudlets in mobile computing. *Pervasive Computing, IEEE*, 8(4):14–23, Oct 2009.

[16] L. A. Setare Amiri, Ahmed Rabbi and R. Fazel-Rezai. *Brain-Computer Interface Systems - Recent Progress and Future Prospects*, chapter A Review of P300, SSVEP, and Hybrid P300/SSVEP Brain- Computer Interface Systems. InTech, 2013.

[17] P. Simoens, T. Verbelen, and B. Dhoedt. Vision: mapping the world in 3d through first-person vision devices with mercator. In *Proceeding of the fourth ACM workshop on Mobile cloud computing and services*, pages 3–8. ACM, 2013.

[18] A. B. Usakli, S. Gurkan, F. Aloise, G. Vecchiato, and F. Babiloni. On the use of electrooculogram for efficient human computer interfaces. *Intell. Neuroscience*, 2010:1:1–1:1, Jan. 2010.

[19] T. Verbelen, P. Simoens, F. De Turck, and B. Dhoedt. Aiolos: Middleware for improving mobile application performance through cyber foraging. *Journal of Systems and Software*, 85(11):2629–2639, 2012.

[20] M. Weiser. The computer for the 21st century. *Scientific american*, 265(3):94–104, 1991.