Interactions with recognized objects

Jacek Ruminski, Adam Bujnowski, Jerzy Wtorek
Department of Biomedical Engineering
FETI, Gdansk University of Technology
Gdansk, Poland
jacek.ruminski@pg.gda.pl

Aliaksei Andrushevich, Martin Biallas, Rolf Kistler
CEESAR-iHomeLab, Technik & Architektur
Hochschule Luzern
Horw, Switzerland
info@iHomeLab.ch

***Abstract*— Implicit interaction combined with object recognition techniques opens a new possibility for gathering data and analyzing user behavior for activity and context recognition. The electronic eyewear platform, eGlasses, is being developed, as an integrated and autonomous system to provide interactions with smart environment. In this paper we present a method for the interactions with the recognized objects that can be used for electronic eyewear. The design of the control node is presented. The control node is the extension of the power socket, equipped with a set of sensors and ZigBee communication interface. The node is used to control electronic objects through user interactions with a mobile device. The preliminary results are presented based on the Android mobile phone used to recognize controllable objects based on the captured image sequences.**

Keywords—human-system interactions; sensor networks; eyewear; smart spaces; intelligent environments

I. INTRODUCTION

The important topic of the interaction with a user in building automation and control is placed under the research area of human-building interaction (HBI). Concretely, new user interface opportunities are being researched actively for the domain of Smart Homes nowadays in order to inform, help, assist and support people in their daily life activities. In summary, Human Building Interaction can be characterized according to several criteria described in [1]:

- interaction type (explicit vs. implicit or invisible) where trade-off between invisibility and added value of HBIs has to be considered to ensure the user acceptance [2, 3];

- abstraction levels of interaction objects (user access to devices, media or services) [4];

- device selection strategy (pre-configured, work as plug'n'play or dynamically orchestrated);

- interaction strategy generation (macros / rules based or by dynamic planning);

- user-related interface modalities (multi-touch control, speech recognition, gesture and pose recognition, user identification and authentication, augmented reality, brain-computer interfaces, etc.).

To ensure the personalized adaptive property of HBI, the system has to rely on the current user context. One implementation challenge relates to context-sharing between users, applications and user interfaces (UI).

The interesting approach for adaptive user interaction with separating the applications from its user interaction is used by open distributed framework of the EU project PERSONA [5]. Using this framework, the applications can be developed completely independent from the input/output devices physically available in the smart environment.

In [6] a portable virtual input control method was proposed for a purpose of human-machine interaction and remote control. The virtual input control device uses an operation interface display unit, an image-capturing module, and a central processing module. The display unit is movably disposed in front of user's head for displaying an operation interface corresponding to a controlled device. The image-capturing module records a position image of a hand of the user. The central processing module transmits display data to the display unit rendering the operation interface. Central processing module also receives the position image of the user's hand captured by the image-capturing module, and determines a control command user input according to the display data and the position image of the user's hand.

Authors in [7] presented the mediated interaction paradigm for gaze-based home automation. In this interaction paradigm, gaze-based actions and reactions are accomplished through a menu-driven control application that allows users to fully interact with the home environment. The scale-invariant feature transform (SIFT) features [8] were used to represent domotic devices. One database is constructed by scanning all the objects and generating the SIFT features. These features are then stored in a database. The Attention Responsive Technology (ART) system [9] constantly monitors the users eye movements and ascertains the allocation of visual attention within the environment, determining whether the users gaze falls on any controllable device. If a device is identified as being gazed at, then the system presents a simple interface dialogue that can be of any form, for instance a touch sensitive screen.

Implicit interaction combined with objects recognition techniques opens a new possibility for gathering data and analyzing user behavior in activity and context recognition tasks. The electronic eyewear platform, eGlasses, is being developed, as an integrated and autonomous system to provide interactions with smart environment. In this paper we present a method and preliminary results for the interactions with the recognized objects that can be used for the electronic eyewear.

978-1-4799-4714-0/14/$31.00 ©2014 IEEE

II. METHOD

A. The interaction method with the recognized objects

The interaction method with the recognized objects assumes that any kind of controllable object (a lamp, a TV set, an air-conditioning system, a robot, etc.) is connected to a control node. The control node is equipped with a wireless interface, a microcontroller, memory and additional sensors or actuators. It is further assumed that developers or manufacturers of the controllable object and/or the control node implemented a software code to control the object and provided the standardized interface for wireless communication. Additionally, it is assumed that the graphical signature of the controllable object is stored in local memory of the control node. The graphical signature of the controllable object is a set of features, which could be used by a mobile device equipped with a camera for automatic object detection using analysis of recorded sequence of images (video). The control node provides a set of control codes that can be requested by the mobile device. The control code is a software library or the identifier that can be used to download a software library. The software library provides functions to read data from the control node (e.g. read a value of a parameter X) or to set a value of a parameter. It could be used to change the state of the controllable object (e.g. turn on the TV, change the temperature, etc.). Additionally, the software library can provide graphical user interface (representation or code) that the mobile device could use for the interaction with the recognized object. The control node could be embedded in the controllable object or it could be a universal, reprogrammable unit.

It is assumed that a user is equipped with a mobile device, preferable electronic eyewear (eGlasses). The mobile device provides a wireless communication interface for the communication with the control node. Additionally, the mobile device uses a camera to record sequences of images from potentially controllable objects in the local surroundings.

The proposed method of interaction between a mobile device and the recognized object is based on two phases: the configuration (discovery) phase and the operational phase. The interaction procedure is illustrated in the Fig. 1.

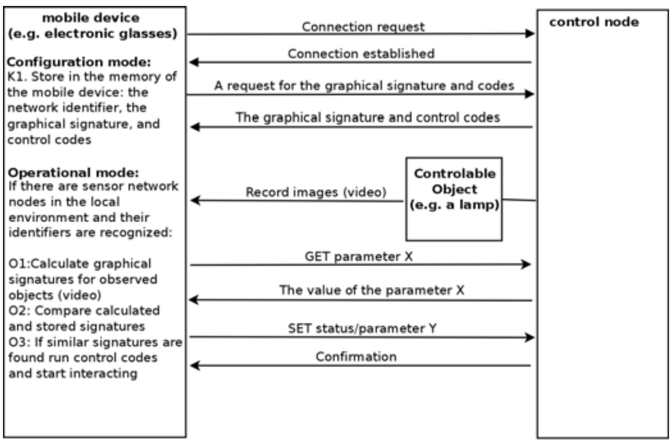

Fig. 1. An illustration of the interaction procedure

The mobile device sends a download request to the control node during the configuration (discovery) phase. In the response The control node provides a data packet containing the graphical signature of the controllable object (the feature vector) and the set of control codes representing the software interface to the connected node. As a result, the representation of the controllable object is stored in the local memory of the mobile device. In the operational phase the mobile device is scanning the local environment. If there are sensor network nodes (control nodes) reachable and recognizable (using data from the configuration phase) in the local network the video recording is starting. Frames of the video are analyzed to detect objects and to generate feature vectors using the same method as has been used to create graphical signatures of controllable objects. In the next step generated feature vectors are compared to graphical signatures stored in the mobile device. Only those stored signatures are used in comparisons, for which the network identifiers have been recognized in the local environment. This contextual information can potentially decrease the total number of comparisons reducing the power consumption of the mobile device. If a similar graphical signature is found in the processed sequence of images then the network identifier assigned for this signature is retrieved from the memory of the mobile device. Additionally, the associated software code (obtained during the configuration phase) is executed. The code provides the interface for interaction with the controllable object (using the control node). The interface can include graphical user interface rendering and a set of functions like: set/get, read/write.

A control node and software for a mobile device have been designed in order to validate the proposed method in terms of interaction with a known object.

B. The control node

The design of the universal control node is based on power socket extension. Two boards have been designed. The board mounted in the lower part of the unit consists mainly of the PIC microcontroller, power/energy meter and audio unit to detect peak values of sound amplitudes. The upper board is equipped mainly with sensors like smoke sensor, humidity sensor, light sensor, microphone as well as with ZigBee wireless communication interface.

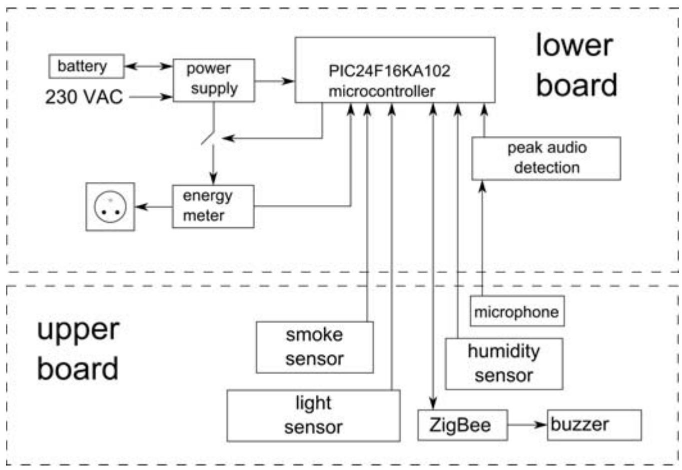

Fig. 2. An illustration of the designed control node

General architecture of the control node is presented in Fig. 2.

Every control node can be uniquely characterized by the network identifier that is represented by a 64-bit MAC address.

C. Object detection methods

It is assumed that at least two categories of methods can be used for the identification of devices: one is based on automatic recognition of object's features and the other is based on visual tags (e.g. Q-R codes). In both examples the graphical signature of the controllable object can be stored as a file in the device or on the network (e.g. implemented as a Web Service). The mobile device can download the file to its local memory using the network identifier. For example, the XML file can be used to store Viola-Jones, Haar cascade data (result of the training provided by the provider of the controllable object) or a set of descriptor values in the metadata file; similar as for the MPEG7 standard.

In the validation reported in this paper we used two simple methods: the method based on Speeded Up Robust Features (SURF features) [10] and the method based on descriptors for shapes and colors. Here, we do not focus on the performance of the object detection methods, but rather on the interaction method.

Finally, the software package has been designed to support wireless communication with the control node, for recognition of objects, and for interaction with objects. Additionally, simple graphical user interface has been designed.

III. RESULTS

The control node and the software for the mobile device have been implemented. The Android 4.2 platform has been used with the support for USB UTG (On-The-Go). The mobile phone has been used for the mobile device role.

A. Communication with the control node

The converter has been developed to bridge USB with simple serial port and thus to provide a communication between the mobile phone and the control node. The ZigBee node has been connected to the mobile phone using the converter. The implemented devices are presented in Fig. 3.

a) b)

Fig. 3. The illustration of the implemented devices: the extended power socket as a control node (a) and the mobile phone connected to the ZigBee module using USB/Serial converter

The simple communication protocol has been implemented based on 4 control messages:

- PINGxxxx control message; where xxxx represents a 4-digits number of milliseconds as interval time for synchronous delivery of ping messages from the control node (for TESTING),
- IGON message – request sent to the control node to turn on the power supply for the controllable object,
- IGOFF message - request sent to the control node to turn off the power supply for the controllable object,
- MEAS message - request sent to the control node to respond with the message containing values measured by sensors of the control node (e.g. power consumption, humidity, etc.).

B. Interaction with the recognized object

A lamp has been used in the experiment as a controllable object connected to the implemented control node (the extended socket). The experiment has been conducted in a home environment choosing different locations of the object and using different light conditions. The illustration of the scanning procedure performed by the implemented as the Android application software module is presented in Fig. 4.

Fig. 4. The illustration of the scanning procedure performed by the software module implemented as the Android application

When the controllable object is recognized the interactive menu with possible actions is rendered. Additionally, the image of the recognized object is shown. The illustration of this step is presented in Fig. 5.

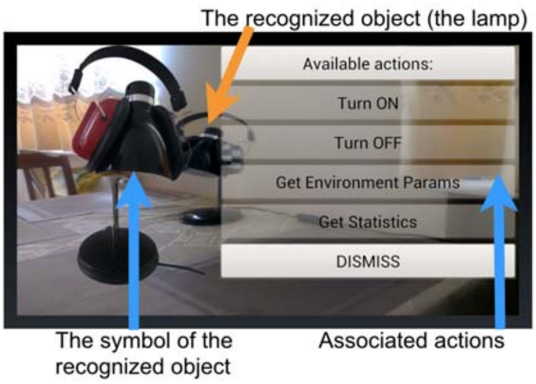

Fig. 5. The controllable object is detected and the menu with available interaction methods is presented

Finally, a user can choose an option from the menu to start the interaction. In the Fig. 6 the case is presented when the user pressed the "Turn ON" button and the light was tuned on.

Fig. 6. The “Turn ON” action was initiated - the light was turned on

The lamp used in the experiment has been characterized by simple graphical representation (simple black shapes). Therefore, the SURF procedure gave poor results in terms of the object detection. The headphones have been added to the lamp to introduce additional, high frequency-based image features for the SURF procedure. This modification gave almost 100% detections, however assuming the same angle of observation and similar lighting conditions. Some examples are presented in Fig. 7.

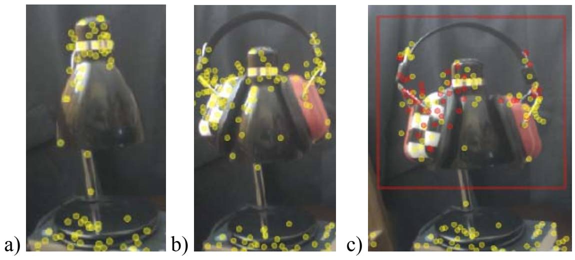

Fig. 7. The SURF features generated for the lamp a), for the lamp with headphones b) and the object detected based on recognized features

IV. Discussion and Conclusions

In this paper we presented a method and preliminary results for the interaction with the recognized objects. The method is based on the recognition of objects using predefined set of features. The set of features is uniquely associated with the network identifier of the node controlling the recognized object.

The control node in the form of extended socket has been designed, implemented and tested. The performed experiments proved that the communication between the mobile device (the Android phone) and the developed node is stable and fast enough to perform real-time interaction. This experiment has been designed and performed to validate the interaction method, which has been designed for the electronic eyewear. The electronic glasses will be used in the eGlasses project (www.e-glasses.info) to control automatically detected, controllable objects. Using the embedded eye-tracker the user will be able to choose the object in the field of view, if there is more than one recognized objects. The idea of the interaction with objects using eyewear is presented in Fig. 8 together with the photography of the first prototype (2010) implemented to test simple eyewear with mobile eye-tracking. In the preliminary testes performed using the setup presented in this paper the simple feature-based object detection methods have been used. Under similar angle of observations and under similar lighting conditions the detection precision has shown itself as satisfactory.

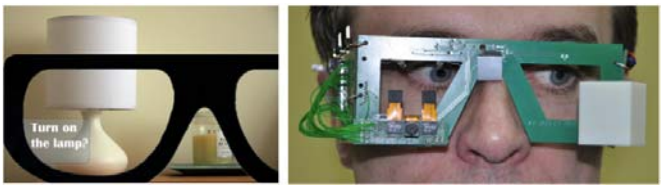

Fig. 8. The idea of interaction with recognised objects using electronic eyewear (eGlasses) – a) and the very early prototype, that was used to verify the mobile tracker activites

However, the results were very poor if the controllable object was rotated, showing the face of the object different than this, for which the graphical signature was created. This problem is very interesting for the further research. The graphical signatures of the controllable objects could include additional identifier, representing the preferred object detection procedure, appropriate for a given vector of features. Additionally, it could be useful to provide not a one graphical signature but a set of signatures (feature vectors), especially for complex and asymmetrical objects. However, this will possibly make the processing tasks more complicated.

Another interesting aspect for the future research is to propose a method for the discrimination of objects that look the same (in sense of graphical signatures) and are spatially located very close to each other. If sensor networks are used, the distance measurement methods based on time-difference-of-arrival (TDOA) [11] can be adopted. Similar method has been presented in [12].

The proposed method for the interaction with recognized objects using electronic eyewear could have many interesting applications. It could be especially useful inside intelligent buildings, e.g. inside living labs like iHomeLab. The introduction of electronic eyewear with interfaces capable to interact with intelligent buildings opens new possibilities in the area of human-system interactions.

Acknowledgment

This work has been partially supported by NCBiR, FWF, SNSF, ANR, and FNR in the framework of the ERA-NET CHIST-ERA II, project *eGLASSES – The interactive eyeglasses for mobile, perceptual computing*, by Statutory Funds of Electronics, Telecommunications and Informatics Faculty, Gdansk University of Technology, and by European Regional Development Fund concerning the project: UDA-POIG.01.03.01-22-139/09-00 - ”Home assistance for elders and disabled -DOMESTIC”, Innovative Economy 2007-2013, National Cohesion Strategy. Additionally we would like to thank Mr. Arkadiusz Palinski for the assistance in the implementation of the extended socket.

References

[1] A. Asghar, N. Shirehjini. “Klassifikation der Human-Environment-Interaction in intelligenten Umgebungen.” In: Hochberger, Christian (Hrsg.) u.a.:Informatik 2006. Informatik für Menschen. Band 2 : Beiträge zur 36. Jahrestagung der Gesellschaft für Informatik e.V.Bonn : Gesellschaft für Informatik, (GI-Edition - Lecture Notes in Informatics (LNI) P-94), pp. 382-389, 2006.

[2] Riva, G., Vatalaro, F., Davide, F., Alcañiz, M.: Interactive Context-Aware Systems Interacting with Ambient Intelligence. In: eds Albrecht Schmidt, IOS Press, 2005

[3] A. Bujnowski, J. Ruminski, A. Palinski, J. Wtorek, "Enhanced remote control providing medical functionalities," *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on* , vol., no., pp.290,293, 5-8 May 2013

[4] Jeff Hsu, Prashanth Mohan, Xiaofan Jiang, Jorge Ortiz, Sushant Shankar, Stephen Dawson-Haggerty, David CullerHBCI: human-building-computer interaction, Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, pp. 55-60, ACM 2010.

[5] Tazari, M.-R.: An Open Distributed Framework for Adaptive User Interaction in Ambient Intelligence., In: Boris E. R. de Ruyter; Reiner Wichert; David V. Keyson; Panos Markopoulos; Norbert Streitz; Monica Divitini; Nikolaos Georgantas & Antonio Mana Gomez, ed., 'AmI' , Springer, , pp. 227-238, (2010)

[6] Ch. Y. Kuo, Ch. L. Chien, Ch. H. Wu, Ch.-W. Chen, Y. K. Tsai, PORTABLE VIRTUAL HUMAN-MACHINE INTERACTION DEVICE AND OPERATION METHOD THEREOF, US Patent, No US 8,555,171 B2, USA, 2013

[7] D. Bonino, E. Castellina, F. Corno, A. Gale, A. Garbo, K. Purdy, F. Shi, A blueprint for integrated eye-controlled environments, Universal Access in the Information Society November 2009, Volume 8, Issue 4, pp 311-321

[8] Lowe D.G. Distinctive Image Features from Scale-Invariant Keypoints. In International Journal of Computer Vision., volume 2, pages 91–110, 2004.

[9] Shi F., Gale A.G., Purdy K.J. Eye-centric ICT control. In Contemporary Ergonomics 2006, (Taylor and Francis, London), pages 215–218, 2006.

[10] SURF-based algoritms developed based on OpenCV library and code provided by Mathieu Labbé, available: https://code.google.com/p/find-object/

[11] M. Pichler, S. Schwarzer, A. Stelzer, M. Vossiek, "Multi-Channel Distance Measurement With IEEE 802.15.4 (ZigBee) Devices," *Selected Topics in Signal Processing, IEEE Journal of* , vol.3, no.5, pp.845,859, Oct. 2009

[12] M. Pezzin, A. Álvarez, "Real life ranging measurements with low power, low data rate transceivers," *Positioning Navigation and Communication (WPNC), 2010 7th Workshop on* , vol., no., pp.247,256, 11-12 March 2010