

---

# Device Recognition for Intuitive Interaction with the Web of Things

**Simon Mayer**  
ETH Zurich  
Universitätstrasse 6  
Zurich, 8092 Switzerland  
simon.mayer@inf.ethz.ch

**Markus Schalch**  
ETH Zurich  
Universitätstrasse 6  
Zurich, 8092 Switzerland  
mschalch@student.ethz.ch

**Marian George**  
ETH Zurich  
Universitätstrasse 6  
Zurich, 8092 Switzerland  
marian.george@inf.ethz.ch

**Gábor Sörös**  
ETH Zurich  
Universitätstrasse 6  
Zurich, 8092 Switzerland  
gabor.soros@inf.ethz.ch

## Abstract

Supporting human users when interacting with smart devices is important to drive the successful adoption of the Internet of Things in people's homes and at their workplaces. In this poster contribution, we present a system that helps users control Web-enabled smart things in their environment. Our approach involves a handheld interaction device that recognizes smart things in its view using state-of-the-art visual object recognition techniques. It then augments the camera feed with appropriate interaction primitives such as knobs or buttons for control, and can also display measured values, for instance, when recognizing a sensor. The interaction primitives are generated from user interface descriptions that are embedded in the Web representations of the smart things. Our prototype implementation achieves frame rates that allow for interactive use of the system by human users, and indeed proved to facilitate the interaction with smart things in a demonstration testbed in our research group.

## Author Keywords

Human-Computer Interaction, Smart Environments, Object Classification, Web of Things.

## ACM Classification Keywords

H.5.3 [Group and Organization Interfaces]: Web-based interaction; I.4.8 [Scene Analysis]: Object recognition.

---

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).  
*UbiComp'13 Adjunct*, September 8–12, 2013, Zurich, Switzerland.  
ACM 978-1-4503-2215-7/13/09.

<http://dx.doi.org/10.1145/2494091.2494168>

## Introduction

In the last few years, an ever increasing number of sensors, actuators, and other everyday objects have been connected to the Internet [6]. This trend particularly encompasses private households [1] where connected household appliances and entertainment devices allow for services such as remote monitoring and control, social media integration, and over-the-air firmware updates. Increasingly being enriched with such “smart things”, households turn into “smart homes” enabling a multitude of novel applications that target domains like energy efficiency, ambient assisted living, entertainment, and security. However, some experts issue warnings about a possible complete loss of control by smart home inhabitants [9], due to the increased difficulty of managing and interacting with their smart devices.

In this poster abstract, we concentrate on the direct interaction of human users with smart things in their surroundings. We describe an approach that combines Web technologies and visual object classification techniques to facilitate the interaction with smart devices and present a prototype implementation of a mobile application that can be used to intuitively interact with smart things.

## System Design

Our system consists of a handheld device (e.g., a tablet computer) that runs an object classification algorithm to categorize items in its view. It contacts recognized smart things to get a description of their user interface in the form of a custom user interface description language. Based on this description, the handheld renders a user interface for recognized devices, thereby transforming the mobile device into a “magic lens” [3] that bridges the real and virtual worlds. On the interface, data provided by the

devices is displayed (e.g., for sensors) and interaction primitives such as buttons or knobs are shown to enable the user to directly interact with them. Figures 1-3 show screenshots of the rendered interfaces for three example scenarios from the home automation domain:

- Interacting with an audio/video system: Using the clickwheel to control the volume (the loudspeaker is used as interaction proxy, Figure 1).
- Interacting with a smart thermostat: Displaying the measured ambient temperature (Figure 2).
- Interaction with a toy robot: Using the graphical buttons to lower and lift the robot arm (Figure 3).

### *Mobile Front-end and Image Recognition*

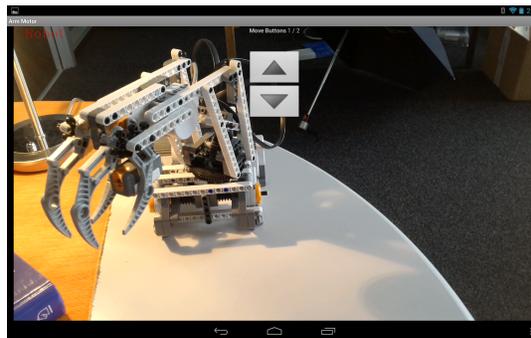
To identify objects in the view of the mobile device, we apply visual classification techniques to extracted local image features that describe the appearance of an object. As features, we chose speeded up robust features (SURF)



**Figure 1:** The handheld device displays a clickwheel-like interface for controlling the volume level of a loudspeaker.



**Figure 2:** Displaying the ambient temperature sensed by a smart thermostat.



**Figure 3:** Controlling the arm of a toy robot.

[2] which are invariant to translation, rotation, scaling, and lighting changes. To be able to use standard classification techniques in our recognition phase, we use the Bag of Words (BoW) [4] model to quantize the extracted SURF descriptors. To classify a given image, we train a linear SVM classifier for each object in our database and take the category that yields the highest classification score. To improve the stability of our recognition results, we use an evidence-based approach

where we report a matching only if at least 5 out of the latest 6 consecutive frames detect the same object.

For the training of our classifiers, we used 10 to 15 training images per smart thing, each representing a different view of the object and extracted 128-dimensional real-valued SURF descriptors from every image. Our system currently can differentiate between eight categories and takes approximately 150ms to process a single frame, thus yielding an interactive framerate of up to 7 frames/s<sup>1</sup>.

#### *Embedded Interface Information*

The smart things we consider include descriptions of an appropriate data type and interaction abstraction to be used for creating interfaces for them as metadata in their Web representations. These descriptions are used to render a suitable user interface, where multiple modalities are supported: Examples of graphical interfaces are shown above, and the system is also capable of rendering speech-based interfaces and of leveraging the gyroscope of our handheld devices (e.g., adjusting the volume by tilting the device). Our interface description information can be embedded in multiple formats, for instance, as JSON documents or as HTML-embedded markup. For this project, we found HTML Microdata to be particularly convenient, because it allows to embed information that is meaningful for humans and machines at the same time (cf. [7] for a more thorough discussion).

#### **Related Work**

The system presented in this paper represents the continuation of earlier work in our group that used 2D tags to identify smart things and then visualized

<sup>1</sup>This was measured on a tablet computer with a 1.7GHz dual-core Cortex-A15 processor and 2 GB of RAM.

interactions *between* these objects on the screen of a mobile device [8]. A project that also targets the augmentation of physical things with graphical user interfaces that are displayed on handheld devices is presented in [5]. While this system allows to render sophisticated 3D scenes on top of physical objects, it also requires 3D models of the devices that may be tedious to create. Our system, in contrast, relies only on snapshots of the objects. Furthermore, it can handle objects that contain rather sparse features, such as the loudspeakers shown in Figure 1, and does not require them to be augmented with distinctive textures.

### Conclusions

We have presented a system that facilitates the interaction of human users with smart devices in their surroundings. Our approach combines live visual object recognition with automatic user interface generation on a handheld device. The handheld is able to identify Web-enabled smart things and to use metadata that is embedded in their Web representations to render a user interface directly on its viewport. Our current prototype achieves frame rates in image processing and interface generation that allow for interactive use of the system while it is robust enough to be applicable in heterogeneous contexts (e.g., regarding lighting conditions or the background of objects). We have successfully tested the prototype application on tablet computers and mobile phones, and plan to extend the system to be able to categorize more devices in the future. Our system cannot distinguish between smart things that look identical. We plan, however, to mitigate this drawback by taking more context information of both the user and the smart things into account (e.g., their location). This could, for instance, allow the system to

differentiate between visually indistinguishable objects that are located in different rooms.

### References

- [1] ABI Research. North America to See Nearly 1.1 Million Managed Home Automation Systems. *Home Networks*, 11(12), 2009.
- [2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.*, 110(3):346–359, June 2008.
- [3] E. A. Bier, M. C. Stone, K. Pier, W. Buxton, and T. D. DeRose. Toolglass and Magic Lenses: The See-through Interface. In *Proc. SIGGRAPH '93*.
- [4] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual Categorization with Bags of Keypoints. In *Workshop on Statistical Learning in Computer Vision (ECCV 2004)*.
- [5] V. Heun, S. Kasahara, and P. Maes. Smarter Objects: Using AR technology to Program Physical Objects and their Interactions. In *CHI Extended Abstracts*, 2013.
- [6] F. Mattern and C. Floerkemeier. From the Internet of Computers to the Internet of Things. In K. Sachs, I. Petrov, and P. Guerrero, editors, *From Active Data Management to Event-Based Systems and More*, volume 6462 of *LNCS*, pages 242–259. Springer, 2010.
- [7] S. Mayer. Service Integration - A Web of Things Perspective. In *W3C Workshop on Data and Services Integration*, Bedford, MA, USA, 2011.
- [8] S. Mayer, C. Beckel, B. Scheidegger, C. Barthels, and G. Sörös. Demo: Uncovering Device Whispers in Smart Homes. In *Proc. MUM 2012*.
- [9] D. Randall. Living Inside a Smart Home: A Case Study. In R. Harper, editor, *Inside the Smart Home*, pages 227–246. Springer, 2003.