Vision-Based Configuration in the Internet of Things: An Example of Connected Lights

Alexander Ilic
University of St. Gallen
St. Gallen, Switzerland
alexander.ilic@unisg.ch

Alexis Leibbrandt
ETH Zurich
Zurich, Switzerland
alexisl@student.ethz.ch

ABSTRACT

Self-configuration and orchestration of networked devices are key areas of the Internet of Things (IoT). In this paper, we explore a prototype system for dynamic configuration that leverages vision-based localization of connected nodes. We study the example of a low-cost and scalable mesh of connected lights that are so seamlessly orchestrated that they can emulate a virtual screen. A smartphone is used for relative localization and control of the mesh nodes. Our findings show how to determine relative positioning of IoT devices with a simple computer vision-based approach and how to synchronize them in a meaningful way in low latency.

Author Keywords

Smart lights; Internet of Things; Mesh Network; Dynamic Configuration; Localization; Mobile; Computer Vision

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; See http://acm.org/about/class/1998 for the full list of ACM classifiers. This section is required.

INTRODUCTION

The Internet of Things describes a vision of how the Internet extends into the physical world [1]–[3]. More and more everyday objects are enhanced with connectivity and thereby made smart. Leading technology firms predict that over 50 billion devices will be connected by 2020 [4].

One of the key challenges is how to make the Internet of Things interoperable and easy to use [5]–[7]. Therefore, it is no surprise that configuration and orchestration of devices is one of the most prolific research areas in the Internet of Things. Several methods [5], [8]–[11] based on semantic models and rule inference has been proposed to enable self-configuration, digital service orchestration, and forming of spontaneous mash ups. Examples span several industries and range from healthcare, to smart grid, and to smart home.

While most of the existing literature is focused on composition of functionality, this paper is looking at the problem of relative localization of each of the connected devices. Typically, low-cost device nodes in the Internet of Things lack the ability to autonomously determine their position in space. Localization with systems such as GPS are too expensive and too coarse in order to establish a meaningful arrangement of nodes. While simple approaches exist to determine proximity (e.g. with signal strength measurements), the problem of determining the relative location of multiple nodes is still not widely addressed. Depending on the application, relative positioning is important for achieving a coordinated goal of meshed devices. In the example of connected lights, localization of the nodes is required to form virtual displays.

In this paper, we develop a low-cost mesh network of connected lights that can configure itself to act as a single composite display. In a novel approach, we use a smartphone to determine the relative position and to control the display of animations.

The paper is structured as follows. In section 2, we outline the requirements, problem statement, and high-level design decisions. In section 3, we introduce our prototype implementation including details on the hardware and software part. We evaluate our solution by simulating typical display tasks from single images up to animations. In section 4, we critically review the limitations of our approach and discuss the implications in section 5. Section 6 concludes the paper and highlights future work.

CONCEPT

Our goal is to demonstrate the value of localization of interconnected devices. We take the example of light devices since these are well suited for demonstrating our concept and quite popular in the smart home area. The solutions of today have already some level of control by smartphones to adjust to a user's context. However, they are often designed as single user or small scale solution. An analysis of existing solutions revealed two key groups of connected lights: small scale home usage (e.g. [12], [13]) and large scale commercial solutions (e.g. [14]).

We aim to combine the best of both worlds by using smart software to control potentially several thousand lights at

IoT'16, November 07-09, 2016, Stuttgart, Germany

DOI: http://dx.doi.org/10.1145/2991561.2991578

low-cost and in a convenient way for any end user. We illustrate the value of device localization by achieving an instant configuration and to simulate with the network of lights the behavior of a single screen.

Requirements and Design Goals

Based on the set goal for achieving a low-cost and scalable mesh network of lights, we focus on the following design goals.

- Each node should be low-cost and flexible in their configuration of light types
- Each node should be able to operate autonomously by itself to facilitate local tasks
- Self-configuration based on semantic profiles should enable the nodes to establish a spontaneous network and assign a master node
- It should be possible to connect to the master node of the network with a standard smartphone for interaction and control
- Each node should allow for connecting different types of light configurations
- Relative localization of the devices should be fast and without user interaction
- At any point, it should be possible to add or remove nodes
- The nodes should act as a single, seamless display
- It should be possible to seamlessly display any image on the composite display
- Animations should be possible in a reasonable framerate

Problem Statement

To address the design goals and requirements above, we pursue the following problem statement:

How we achieve a low-cost, dynamically scalable local network of interconnected lights that is usable as a single screen by a regular user.

Approach

To achieve the low-cost approach, we do not add any self-localization to the nodes. Instead, we use the smartphone's camera to localize the mesh nodes relatively to each other and to determine their orientation with a computer vision approach.

To achieve a low-cost hardware set-up, the biggest design decision is the choice of the wireless connectivity. On the one hand, it should be simple for a user to connect and on the other hand it should be high performance for controlling the whole network.

One of the first popular solutions was the Philips Hue [12] which is based on a Zigbee network and requires a separate WiFi bridge to allow control by a smartphone. According to the vendor it supports up to 50 lights. A more advanced approaches is done by LIFX [13] which does not need an additional bridge. The mesh network is based on 802.15.4 6LoWPAN and can connect up to 100 lights according to the vendor. Automatically, one of the lights configures itself as a master which provides a WiFi control interface. Unfortunately, both solutions are relatively costly and not designed for high scalability.

After studying these two solutions, we also evaluated other suitable wireless protocols [15], [16]. We selected ANT in a practical evaluation for several reasons. ANT is a protocol that is typically used in sports and fitness applications. It operates in the 2.4 GHz frequencies and is particularly well suited for scalable networks. The topology of the network can even be reconfigured at runtime. The nodes of an ANT network communicate through channels. One node can have many channels. It can for example be master of a channel and slave of another of its channels. The communication is typically sequential. This means that the master sends a packet of data (typically 8 Bytes) at a given period continuously. The transmission can be unidirectional without any response from the slave (broadcast mode) or bidirectional with an acknowledgement from the slave for critical data transmission (acknowledged mode). If more data throughput is required, a node can at any time send burst data (standard burst up to 20kbps). Depending on the configuration of the channel, a single master can reach many slaves.

ANT also offers other interesting features. An asynchronous mode can be used to send data asynchronously, for example when a specific event occurs. A background scanning mode enables slaves to listen for available masters continuously without being explicitly connected to these masters. A last mechanism allows for a slave node to be paired with a master node event without knowing the detailed configuration of the master. This can also be done based on the relative proximity of the devices.

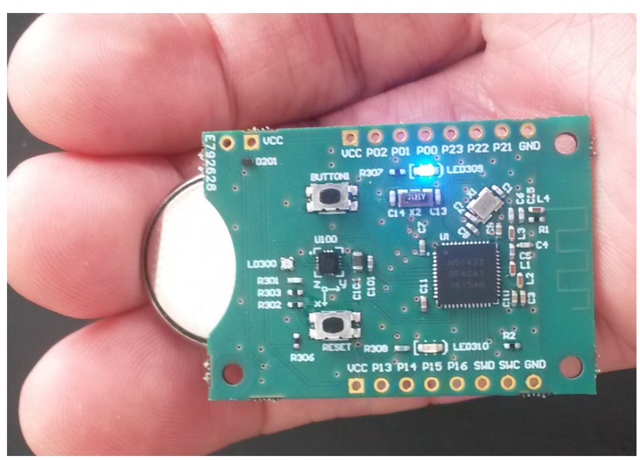

Fig. 1. Final iteration of prototype node

Fig. 2. The boards are packaged in a wooden box and LEDs are behind a semi-transparent screen. Top image shows in off state and bottom image with a few random LEDs in on state.

Based on these considerations, we found that the cheapest combination that matches our requirements is a combination of Bluetooth Low Energy as control interface and ANT for the light network.

PROTOTYPE IMPLEMENTATION AND EVALUATION

We have implemented the project over a time of one year and created several iterations of the prototype. The final prototype board used in this paper is shown on Fig. 1. The board has a size of 48mm x 22mm and can be powered by a standard button cell. In addition, it also supports an additional rechargeable battery to power more peripherals and can be charged via USB. The board bases on a dual radio nRF51422 chip by Nordic based on a 32-bit ARM Cortex M0 core with 256kB/128kB flash and 32kB/16kB RAM. The dual radio features the simultaneous usage of Bluetooth Low Energy (BLE) and the ANT protocol for mesh networking. We designed the board ourselves and included several configurable digital and analog inputs and outputs. Also, we included an inertial measurement unit (IMU) to determine orientation and acceleration of the board.

As seen on Fig. 2, we embedded the board in a wooden box with a simple semi-transparent screen, and connected it to an array of 64 Neopixel LEDs. The LEDs support 24-bit color and are arranged in 8x8 layout.

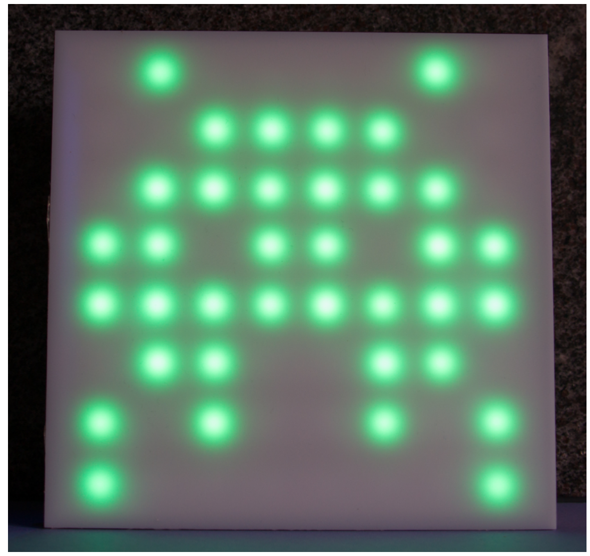

Fig. 3. Visualization of an image on a single node

Experimental Set-up

The ANT nodes are configured to dynamically form a mesh network once they are powered on. Currently, we use a tree hierarchy. One of the nodes configures itself to act as a master and provides also a Bluetooth BLE interface. With a smartphone app for Android, we can connect to the master node and thereby control and interact with the network. Each node has a connected light output and can also display an image completely autonomously (Fig. 3). The light output can be either directly powered from the device or use an additional battery. In practice, we can use completely heterogeneous lamps (e.g. LEDs, light bulbs, etc.) connected to each node. We have already envisioned in the self-descriptor of each node an abstraction that only reports the display size that could range from 1x1 to 1024x1024 and come in any rectangular shape. For illustration purposes, we connected each of the nodes with an 8x8 matrix of LEDs.

Relative Node Localization

After the enumeration and fusion of the semantic descriptors of the nodes, we want to determine the relative position of each of the nodes. The user simply points the

Fig. 4. Prototype set-up showing three nodes with a random pattern and a smartphone for doing the camera-based localization

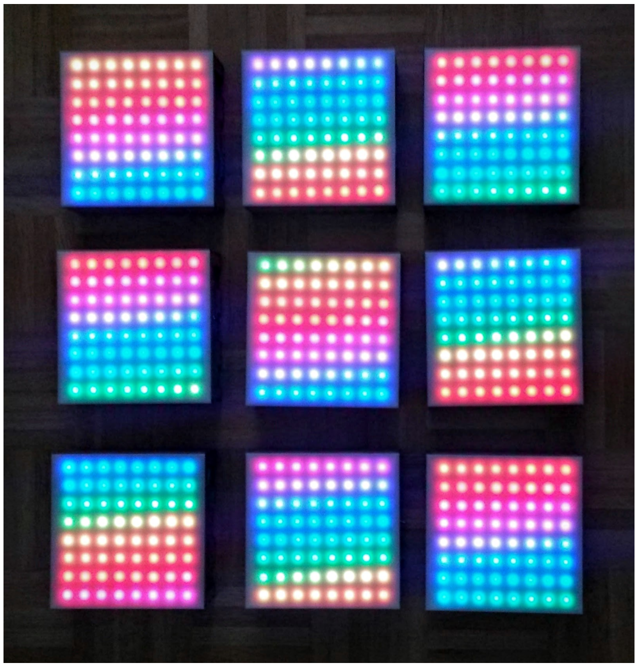

Fig. 5. Nine nodes in random local pattern

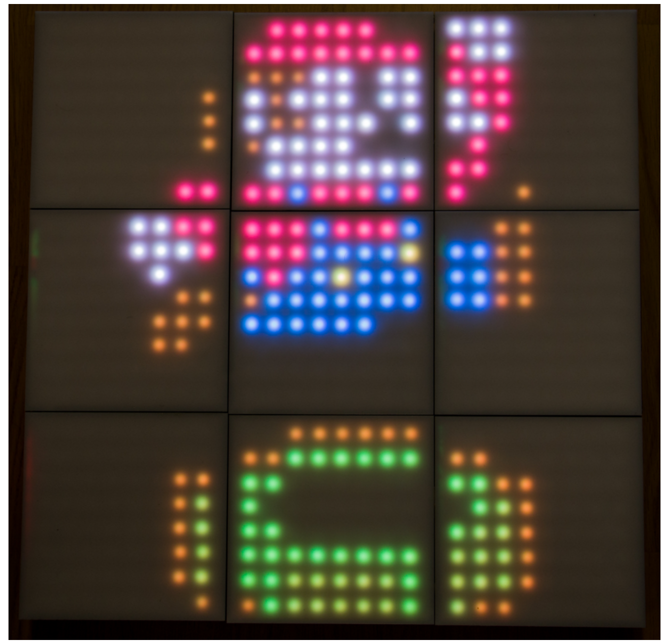

Fig. 6. Evaluation result shows nine nodes behaving as a single display

smartphone to the nodes that should be configured (Fig. 4). In configuration mode, the Android app sends known, random patterns to each of the nodes via the master node. We use the IMU of each node to determine the orientation of the expected rectangular grids. We use a simple OpenCV implementation based on standard feature descriptors and homography-based matching to determine the pose of each of the nodes in 3D space and to compensate for perspective distortion. Based on the camera image and the matched patterns, we derive a rectangular, composite display matrix. If all nodes of the mesh network are visible on the camera, we determine their relative position and fit them into the composite grid. Please note that in this approach, also spots of the grid can remain empty. To avoid issues with moving objects, we imply a static scene assumption and only consider nodes with zero acceleration of the IMU. Also, we invalidate nodes if acceleration is detected to highlight that a recalibration is needed.

Orchestrating the Display

One of our goals was that the nodes are so well orchestrated that they behave like a single screen (Fig. 5). To demonstrate this capability, we use animations that span several frames and several nodes.

With the smartphone app, we can choose arbitrary pictures as key frames. The pictures are scaled down to the virtual display size and receive a frame number. Each frame is then segmented and transferred via Bluetooth BLE to the master node. The master node pushes then the frames to the rest of the network with the ANT protocol.

Initially, we used the maximum ANT message rate of 200 Hz with 8 Bytes of data payload to communicate from one node to another. Assuming that we want to support simple animations with at least five image switches per second with nodes of 64 pixels, we can clearly see that this is not enough if we have more than one node. In a naive encoding, we need at least 3 Bytes for color (RGB) per pixel and an additional Byte for timestamp and control data.

To overcome this limitation, we use the burst mode to distribute the image data. The burst mode achieves an up to 4.6 times higher transfer rate but requires significantly more power. In our setup, the time to distribute a single frame into all of the nodes was around 900ms.

Once the images are distributed to each of the corresponding nodes, we use a simple broadcast command to tell all nodes synchronously to switch to a certain pre-loaded image (Fig. 6). For the animations, we benefit from the synchronization between master and slaves from the ANT communication. After receiving the general broadcast command, all nodes of the system loop trough their respective stored images in a perfectly synchronized way taking as reference the synchronization mechanism of the ANT protocol. By doing so, we can achieve a continuous, high performance animation in a reliable way.

LIMITATIONS

Our approach is a first prototype and therefore has several limitations which we discuss below.

First, the resolution of our approach is limited when compared to a typical smartphone display. In our current prototype, we only use 576 pixels instead of a phone with over one million pixels. Theoretically, already our current setup could be extended to support 4 megapixel (up to 65'000 nodes due to ANT limitation). However, even with a small number of pixels there is the ability to do impressive things. For instance MIT students turned a whole building into a Tetris game with 153 pixels [17].

Second, the distribution time for a single image to the whole network takes currently around 1 second. This is not yet practical for large scale networks. Since we only used a simple approach without compression, this can be further improved in the future.

Third, our localization approach currently requires that all nodes are visible on the same camera image. Also, the distance from smartphone to nodes for localization is currently limited. While the radio protocols for BLE/ANT could in theory support longer distances (e.g. > 20 meters), we only verified the localization in a few meters range. To fully leverage a more wide-spread area of nodes, the camera resolution would have to be increased and or complemented with more advanced computer vision methods for positional tracking.

Fourth, while our BLE/ANT-based solution was very robust and provided a low latency experience, we did not explore the impact of interference in a busy environment and with a significantly larger node size.

Finally, we currently assume a static scene which makes it not suitable to support cases with dynamic movements. For example, there are some high-end connected light solutions that support people to become pixels with wristbands [14]. To support this dynamic mode with moving pixels, a smarter localization and a continuous tracking algorithm could be a suitable way to overcome this.

DISCUSSION

Until now, location information of connected devices was mostly hardcoded in semantic configuration profiles or provided by the users. We envision a next step of connected devices which can dynamically determine their position. Methods might include either a built-in localization such as triangulation with radio signals or visual-based localization. As shown in this paper, localization can be achieved without adding sensors to the individual nodes and thus provided in a scalable way by just using the user's smartphone camera.

Future work should include more advanced method for dealing with dynamic scenes and also for a quicker detection of the different node position by using a parallel pattern detection approach. Also, we plan to extend the prototype trial to the next order of magnitude of devices. In the present case, we have already validated our approach with about ten times the connected amount of LEDs that current commercial smart light solutions in the low price segment support. Our goal is to achieve the barrier of extending to several thousand node points for which the self-configuration and localization will be even more important.

In addition to enabling new functionality such as composite virtual displays as outlined in this paper, this would also allow for more natural interaction with other modalities (e.g. "turn on the light on the left side").

Computer vision methods have been so far only rarely used in combination with Internet of Things. Cameras might develop themselves towards the most important sensors. Not only because they are part of smartphones, our personal gateways to IoT, but also because they are accessible at a very low price point so that future solutions might even include a camera on each connected devices for "inside-out" localization capabilities. The merge between computer vision and Internet of Things applications will be further accelerated by technology trends such as Augmented Reality where the visual identification and interaction might be the most important and natural way to orchestrate actions across multiple devices.

CONCLUSION

As one of the first papers, we studied in additional to automatic configuration on a functional level the role of localization for IoT devices. Our prototype is based on a low-cost design with no localization capabilities provided by the individual nodes (i.e. no GPS, no cameras). Instead, we showed how an automatically, self-configuring mesh network can be enhanced with node localization information with a standard smartphone. We evaluated our solution in an example application of a virtual display to display pictures and even animations. Our solution provides an interesting pathway for localizing objects in the smart home to provide better context and new functionality to users. In addition, we see further changes to advance the research towards Augmented Reality to orchestrate devices on a new level.

REFERENCES

1. D. Uckelmann, M. Harrison, and F. Michahelles, "Architecting the Internet of Things," *Archit. Internet Things*, no. JANUARY, pp. 1–353, 2011.
2. D. Guinard, V. Trifa, F. Mattern, and E. Wilde, *Architecting the Internet of Things*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
3. M. Köhler, D. Wörner, and F. Wortmann, "Platforms for the Internet of Things – An Analysis of Existing Solutions," *5th Bosch Conf. Syst. Softw. Eng.*, 2014.
4. Cisco, "The Internet of Things." [Online]. Available: http://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf. [Accessed: 06-Jun-2016].
5. C. Lerche, K. Hartke, and M. Kovatsch, "Industry adoption of the Internet of Things: A constrained application protocol survey," *IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA*, 2012.

6. M. Elkhodr, S. Shahrestani, and H. Cheung, "The Internet of Things : New Interoperability, Management and Security Challenges," *Int. J. Netw. Secur. Its Appl.*, vol. 8, no. 2, pp. 85–102, 2016.
7. G. Aloi, G. Caliciuri, G. Fortino, R. Gravina, P. Pace, W. Russo, and C. Savaglio, "A Mobile Multi-Technology Gateway to Enable IoT Interoperability," *2016 IEEE First Int. Conf. Internet-of-Things Des. Implement.*, pp. 259–264, 2016.
8. S. Mayer, R. Verborgh, M. Kovatsch, and F. Mattern, "Smart Configuration of Smart Environments," *IEEE Trans. Autom. Sci. Eng.*, vol. PP, no. 99, pp. 1–9, 2016.
9. P. States, D. Mashups, O. Questions, M. Kovatsch, and S. Mayer, "Practical Semantics for the Internet of Things :," *Internet of Things (IOT)*, 2015.
10. D. Guinard and V. Trifa, "Towards the Web of Things : Web Mashups for Embedded Devices," in *Proceedings of WWW*, 2009.
11. R. Kleinfeld, S. Steglich, L. Radziwonowicz, and C. Doukas, "Glue.Things: A Mashup Platform for Wiring the Internet of Things with the Internet of Services," *Proc. 5th Int. Work. Web Things*, pp. 16–21, 2014.
12. Philips, "Hue." [Online]. Available: http://www.meethue.com. [Accessed: 06-Jun-2016].
13. LIFX, "LIFX." [Online]. Available: http://www.lifx.com. [Accessed: 06-Jun-2016].
14. Glowmotion, "Wireless LED Wristbands." [Online]. Available: http://www.glowmotiontechnologies.com. [Accessed: 06-Jun-2016].
15. A. Dementyev, S. Hodges, S. Taylor, and J. Smith, "Power consumption analysis of Bluetooth Low Energy, ZigBee and ANT sensor nodes in a cyclic sleep scenario," *2013 IEEE Int. Wirel. Symp. IWS 2013*, pp. 2–5, 2013.
16. M. Chen, S. Gonzalez, A. Vasilakos, H. Cao, and V. C. M. Leung, "Body area networks: A survey," *Mob. Networks Appl.*, vol. 16, no. 2, pp. 171–193, 2011.
17. MIT, "Tetris on the Green Building." [Online]. Available: http://hacks.mit.edu/by_year/2012/tetris/. [Accessed: 06-Jun-2016].