

Development of a Light-Tracking and -Redirecting System Actuated by Hand-Gesture Recognition

Alexander Liu Cheng^{1,2}, Nestor Llorca Vega^{2,3}, Galoget Latorre⁴, Daniel Coba²

¹Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, The Netherlands

²Facultad de Arquitectura e Ingenierías, Universidad Internacional SEK, Quito, Ecuador

³Escuela de Arquitectura, Universidad de Alcalá, Madrid, Spain

⁴Centro de Educación Continua, Escuela Politécnica Nacional, Quito, Ecuador

E-mail: a.liucheng@tudelft.nl, nestor.llorca.arq@uisek.edu.ec, galoget.latorre@epn.edu.ec, dcoba.arq@uisek.edu.ec

Abstract—This paper presents a context-aware light-tracking and -redirecting system guided by hand-gesture recognition. It is conceived as yet another mechanism within an ongoing development of a more intuitive and technically sophisticated Ambient Intelligence / Active and Assisted Living ecosystem. The detailed system consists of individual nodes that are strategically installed across regions of a building-envelope, which enables this latter to draw or deflect direct natural light into or away from specific locations within the built-environment as requested by the user(s) via recognized hand-gestures. Each node is capable of sending and receiving sensed-data continuously via ZigBee with one another as well as with microcontrollers embedded within the interior built-environment. Said microcontrollers are equipped with cameras via which four hand-gestures may be recognized. The first or initializing hand-gesture engages the system and enables it to recognize any of the remaining hand-gestures. The second redirects light towards the position of the detected hand-gesture, while the third redirects it away from said position. Finally, the fourth gesture turns the light-tracking and -redirecting system off.

Index Terms—Internet of Things, Wireless Sensor Networks, Ambient Intelligence, Active and Assisted Living, Adaptive Architecture

I. INTRODUCTION

The present light-tracking and -redirecting system is built upon an ongoing development of a highly scalable and technically robust *Ambient Intelligence* (AmI) / *Active and Assisted Living* (AAL) System Architecture [1]. It is conceived as another service feature in a growing list of features enabled by said System Architecture. These features are geared towards the enhancement of quality of life with respect to *Activities of Daily Living* (ADLs) as well as with respect to spatial experience within the built-environment.

The objective of the system described in the present paper is to enable the user(s) to draw or deflect direct natural light into or away from his/her inhabited space in a way that enhances its quality (for examples of notable built-environments enhanced by direct natural lighting considerations, see Steane [2]). This notion of quality labors on subjective considerations such as human preference as

well as objective ones such as measurable health benefits associated with exposure to direct natural light. For example, humans prefer direct natural light over artificial counterparts (especially fluorescent) as they associate it with higher levels of concentration, health, aesthetic quality [3]. Additionally, daylight is associated with fatigue mitigation as well as vitality, which is a psychological metric for mental well-being [4, 5]. Artificial lighting, especially in the workplace and in the pervasive form of fluorescent lights, is associated with fatigue and degraded cognitive performance [6]. Furthermore, lack of exposure to direct natural light is correlated with depression and even to the perception of sleep quality, where insufficient exposure to direct natural light is strongly correlated with the perception of insufficient sleep [7]. The character of the direct natural light drawn by the nodes of this system is diffused and not focused, and its diffused in such a way as to bring daylight qualities into the built-environment and not as to direct a spotlight on a particular region, which would cause discomfort and be potentially detrimental to the user(s)'s well-being. This system is conceived as a support layer for a class of dynamic building-envelopes—capable of reacting to a particular set of environment-related conditions—such as Jean Nouvel's dilating façade components in his *Institut du Monde Arabe* [8], Aedas®'s *Al Bahar Towers* [9], ETH Zürich's *Adaptive solar façade* [10], and the adaptive and context-aware building-envelope nodes previously developed by the authors [11] to name a few. The present system is conceived and construed as an active protagonist in the continuous promotion of the user(s)'s well-being and spatial experience, and it is considered as part of an intelligent built-environment ecosystem centered on Internet of Things and People, or what Oosterhuis has called a Society of Home [12] in describing the future dwelling space.

The system is detailed and discussed in the following four sections. Section II describes the overall concept of the system. Section III details its implementation methodology in the present *proof-of-concept* setup. Finally, Section IV concludes with a discussion on limitations and future work.

II. CONCEPT

The system enables the user(s) to draw or deflect direct natural light via a series of discrete nodes equipped with concave mirrors whose position reconfigures horizontally and vertically (see Figure 1). These nodes are intended to be integrated in specific regions of the building envelope, depending on the orientation of daylight sources given the particular geographic location on which the built-environment is situated. It is not necessary to populate the entire building-envelope with these nodes given that certain latitudes do not receive direct natural light on certain facades throughout the year. Said nodes continuously send and receive information with microcontrollers (MCUs) embedded in the interior space that serve a variety of other service features inherited from previous work. Some of these MCUs, which are strategically installed in specific locations within the interior space, are equipped with low-cost yet effective cameras (i.e., *Raspberry Pi Camera v2*) capable of engaging in computer vision functions such as facial-recognition [13] as well as gesture-recognition (see Section 0). A user in the interior of a built-environment covered by the service range of a given set of nodes (again, depending on where the nodes are strategically installed considering the building's geolocation) engages the light-tracking and -redirecting system by effecting an initializing hand-gesture (viz., gesture 1; see Figure 3, *Top-left*). Due to the position of the cameras installed across said space, the general position of the recognized hand-gesture may be ascertained. It is in this manner that the nodes may reconfigure to redirect their received direct natural light towards the general position of a detected specific hand-gesture whenever this is recognized (viz., gesture 2; see Figure 3, *Top-right*). It should be noted that if the nodes that service the interior space in question are not receiving direct natural light and are therefore unable to redirect it, a sound-notification is emitted to inform the user. In addition to drawing direct natural light, the nodes may also deflect it from a hand-gesture location when another specific hand-gesture is recognized (viz., gesture 3; see Figure 3, *Bottom-left*) and if its concave mirror's vertical and horizontal range is within the direct natural light's direction. Since each discrete node is integrated with four *Light-Dependent Resistors* (LDRs), it is capable of ascertaining whether it is in the path of direct natural light. If after being engaged by gesture 3 none of the nodes servicing a given interior space detects the presence of direct natural light within its vertical and horizontal displacement—and can therefore not engage in deflection / redirection—a sound-notification is emitted to inform the user. Finally, when gesture 4 (see Figure 3, *Bottom-right*) is recognized, the light-tracking and -redirecting system is instructed to stop listening to gestures 2 and 3. As the cameras in the built-environment are continuously engaging in object, facial-identity and -expression, and hand-gesture recognition, this last gesture is implemented in order to prevent the false engagement of the nodes via gestures 2 and 3.

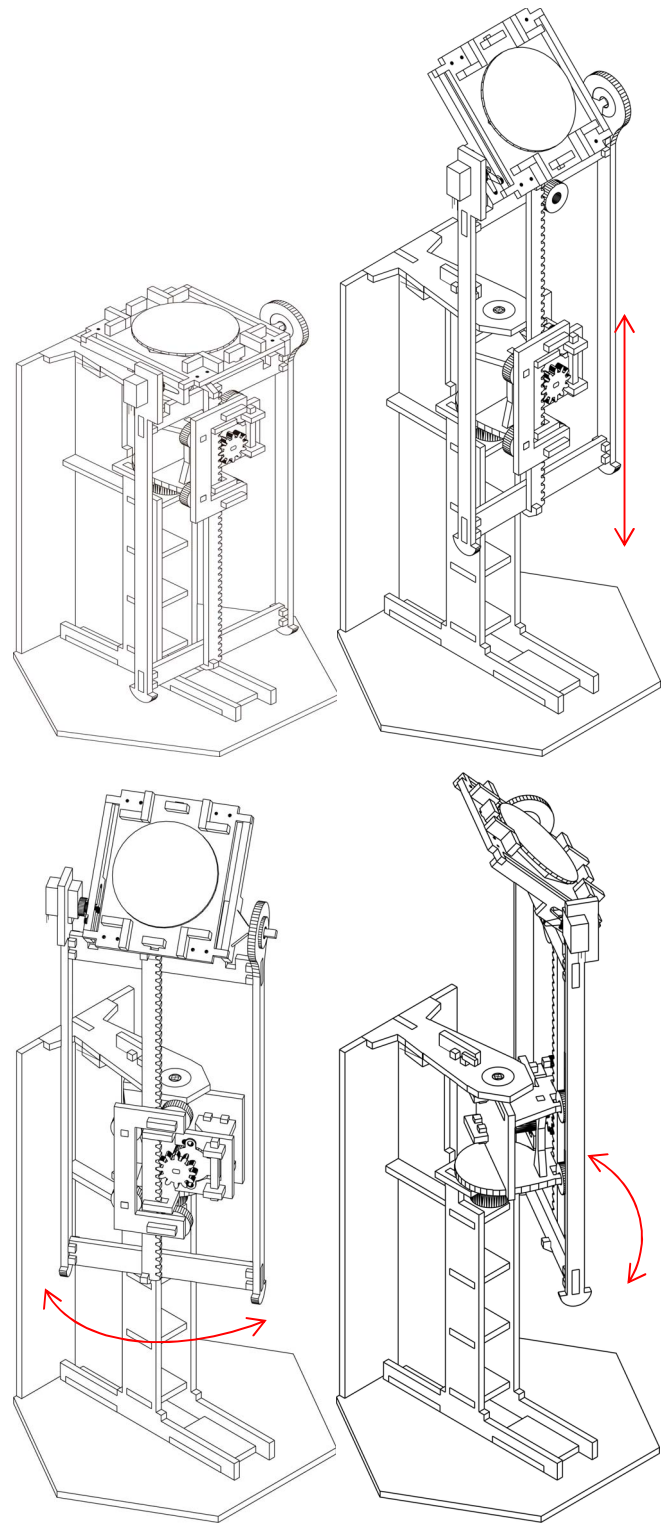


Figure 1. Top-left: Initial position of light-tracking module. Top-right: Vertical adaptation of module. Bottom-left: Module turning right. Bottom-right: Module turning left.

III. METHODOLOGY

The present light-tracking and -redirecting system consists of two main components: (1) a set of discrete light-tracking and -redirecting nodes, and (2) a hand-gesture mechanism that commands them. Although these are integrated to work as a single system, the present section discusses the implementation of each component separately.

A. Light-tracking Modules and system

Although each node is its own discrete mechanism, the present setup has one microcontroller (MCU) for every set of three nodes, which occupy all of the MCU's analog and digital GPIOs. A single *XBee Series 2 antenna* services each MCU. Nevertheless, the individual sensor data of each node may be identified via software. That is, the perceived light-intensity data of each node is stored in an array, and the source node of a given data stream may be identified via the index values of this array. In this manner any node from within any cluster of three (see Figure 2, *Top*) may access the sensed-data of any other node within any other cluster of three. There are scenarios where such interconnectivity is not necessary in the present setup. For example, in order to orient a particular node's mirror a particular location, the node only needs to receive the instructions from its MCU. Nevertheless, envisioned future scenarios and services require the nodes to be capable of swarm-like behavior, which is effectively only possible within the context of a meshed topology.

Each node consists of two stepper motors for vertical displacement and horizontal rotation as well as one servo motor for the vertical rotation of the concave mirror (see Figure 2). The surface that supports the concave mirror is integrated with four LDRs, and the direction of said surface's displacement is determined by which LDR or combination of LDRs is receiving the most intensity of direct natural light. Based on this information, the MCU in conjunction with the camera data sent to it by the camera-integrated interior MCUs computes the horizontal and vertical displacements as well as vertical rotation necessary to redirect sensed direct natural light towards a specific location within the built-environment (as guided by gesture 2). A similar but opposite computation takes place in order to deflect detected direct natural light from a given space. In the present setup the nodes are configured to deflect to any direction that requires the least amount of physical displacement as long as no direct light is detected at the location of the hand effecting gesture 3. In subsequent implementations a default location may be identified for mirrors to focus on while deflecting light from a given location. But in the present *proof-of-concept*, the least amount of physical displacement is informed by energy-efficiency considerations.

In order to preserve energy, the MCU controlling each cluster sets each of its nodes on standby and the XBee antenna to sleep-mode when no direct natural light is detected.

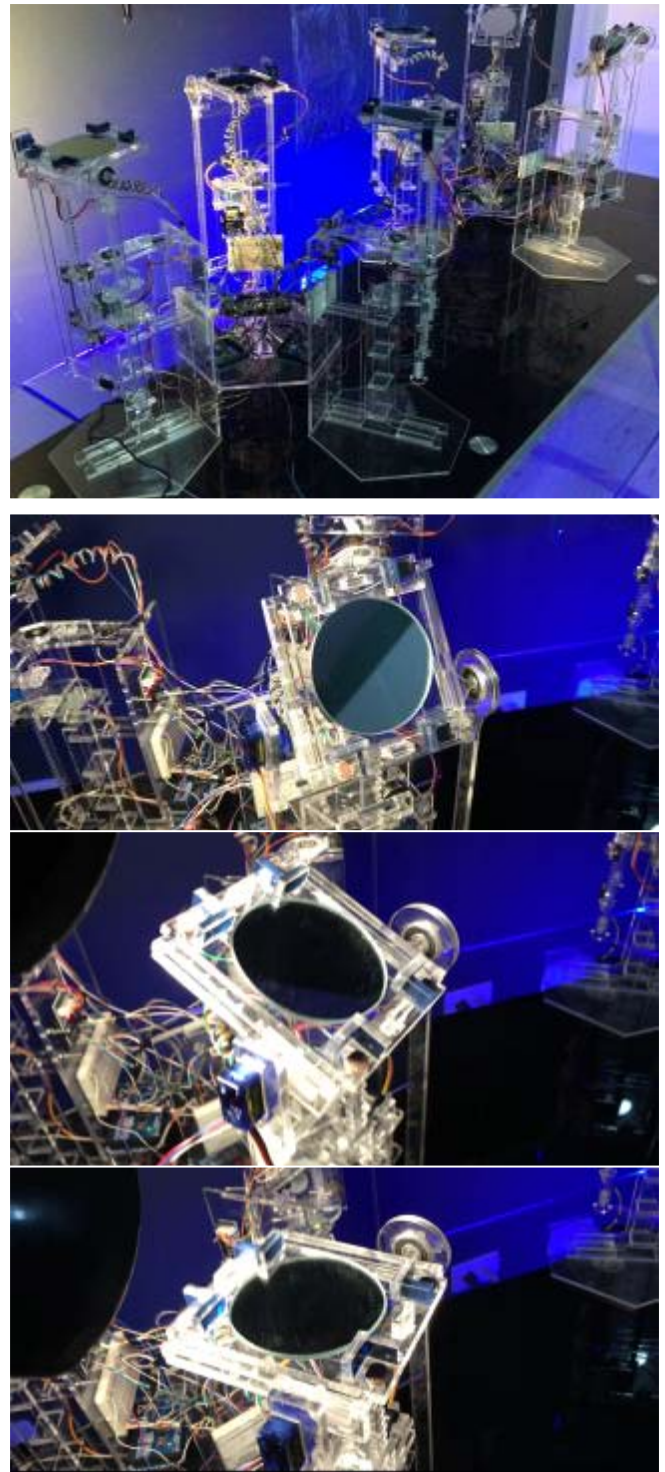


Figure 2. Top: cluster of three nodes. Second image to bottom: mirror surface's vertical rotation sequence.

B. Hand-Gesture Recognition Mechanism

As mentioned in Section I, the present implementation inherits and builds upon the open-ended System Architecture previously developed by the authors. In particular, (1) a visual gesture-detection mechanism is built next to an inherited object-recognition mechanism [14] in order to

enable the intelligent built-environment to be capable of hand-based gesture-recognition; and (2) an extension of an inherited *Human Activity Recognition* (HAR) mechanism is used to make inferences about motion and basic posture with respect to the user. This same mechanism has previously been used to trigger actuation events such as ventilation, illumination, etc., within the built-environment. Nevertheless, the present discussion is limited to the light-tracking and -redirecting system in question. The visual gesture-detection mechanism relies primarily on architecture-embedded ICTs while the HAR mechanism—and therefore the motion and basic posture recognition components thereof—relies primarily on smart-devices and wearables (which transmit body-specific accelerometer and gyroscopic data) and secondarily on embedded ICTs (which receive and process said data).

The instantiated feature in question enables the user to trigger physical and computational events by a combination of recognized hand-gestures in relation to recognized posture and/or motion associated with specific activities. In order to provide a technical breakdown of the mechanisms that enable the feature in question, this section details the implementation of the hand-gesture mechanism first and the extension of the HAR mechanism second, culminating with a description of how their outputs are used in conjunction to trigger events. The inherited object-recognition mechanism [14] is built with BerryNet® [15], which in turn is built with Inception® ver. 3 [16] for a classification model and TinyYOLO® [17] for a detection model. A salient advantage of BerryNet is that it is a fully localized gateway implementable on a cluster of Raspberry Pi 3s (RPI3s), which are the principal computational nodes of the inherited System Architecture. On an individual RPI3, the inference process is slow, requiring a delay between object-recognition sessions, which is ameliorated by the dynamic clustering feature of the inherited *Wireless Sensor and Actuator Network* (WSAN) that supervenes on said System Architecture. Incidentally, it is worth noting that due to their evolving and resilient characters, *Machine Learning* (ML) classifiers are aptly implemented in a wide-range of applications built on WSANs [18]. Another benefit-cum-limitation is that BerryNet®'s classification and detection models are pretrained, which avoids the need to generate said models locally, but which also restricts its range of detection. However, for the present implementation this is more a benefit than a limitation, since the object-detection mechanism is used to recognize general objects (e.g., cars, chairs, people, dogs, etc.) but not to discern between variations or specifics. Within the intelligent built-environment, a number of cameras and corresponding RPI3s embedded within deliberately purpose-fabricated cavities are used via the object-recognition mechanism to recognize whether a person is within a given area of the built-environment. The hand-gesture mechanism is executed only when a person is detected with a fair degree of confidence (i.e., confidence-level greater than 75% in BerryNet's inference results). This is informed by efficiency

considerations, as there is no need recognize for hand-gestures where there is an absence of hands. In addition to saving energy and computation resources, this consideration also decreases the number of false positives in improbable-yet-possible scenarios. For example, a computer screen in a given space may display an image of a hand in a recognizable gesture, which could trigger the hand-gesture mechanism to construe a detected instance when there is none. Also, there may be a physical model of a hand bearing a particular gesture—e.g., a mannequin's hand—which may also fool the hand-gesture mechanism. These two scenarios, however improbable, are wholly avoided by enabling the object-recognition mechanism to allow the hand-gesture recognition mechanism to consider only those instances where a person's presence has already been confirmed. A caveat: there are measures that could be implemented to discard displayed hands or physical representations of hands such as using sensors to gauge the heat-signature and -distribution of the object, but even with such measures there would be a question of how efficient it would be to run both object-recognition and hand-gesture recognition mechanisms simultaneously. To be sure, there may be instances where such operation is justified, but such considerations exceed the scope of the present implementation.

The hand-gesture mechanism is built with OpenCV [19]. Unlike the object-recognition mechanism, it is deployed locally in its entirety due to its relatively modest consumption of computational resources. In this setup, the mechanism first generates a silhouette corresponding to the person detected by the object-recognition mechanism. It then identifies the regions where the hands should be and centers on capturing images of those regions. A rough rectangular region is drawn around the object captured in said images and its contours are identified via segmentation techniques based on the thresholding of images in grayscale. This results in a boundary line representative of the captured object (see Figure 3). Note that at this point the object may not actually be a hand. There may have been errors in the process of detecting the regions where hands are typically found, especially in non-controlled setups (see Section IV for a brief discussion on the limitations of transplanting the present work into non-controlled environments). Accordingly, further steps are taken to ascertain whether it is a hand bearing any one of the three considered gestures. The first is to identify the tips of the objects that may or may not be fingers as well as the concavities between them. From these points angles are calculated, and if they are within specified limits, then a high-probability of it being a hand via identification of fingers is construed. In this same step the number of fingers is identified and recognized as any of the three considered gestures. It should be noted that segmentation techniques presently employed will not work in all scenarios, especially in when illumination and composition are not controlled. In such cases other segmentation techniques may be applied—for example, Soffritti uses skin tone as a threshold to binarize images [20].

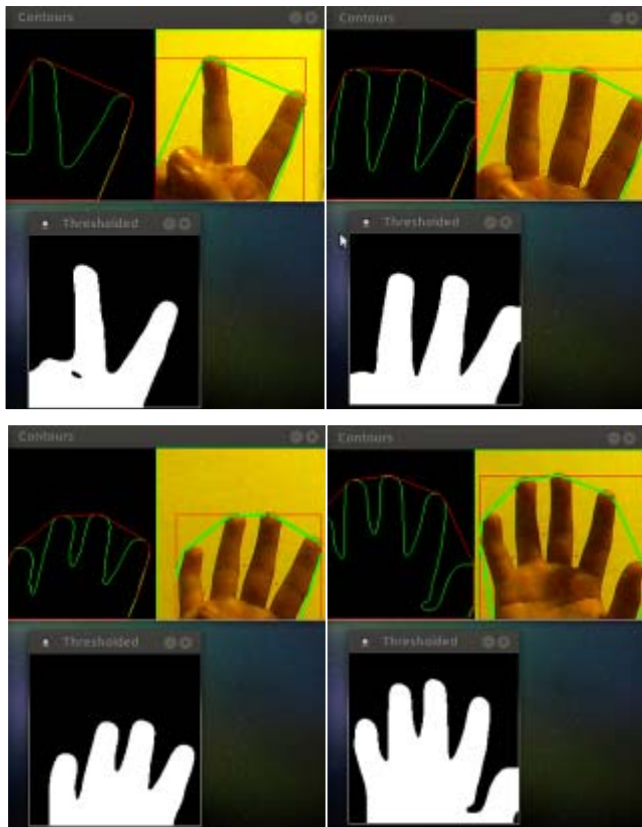


Figure 3. Top-left: Gesture 1: System Initialization. Top-right: Gesture 2: Draw light towards the location of hand-gesture. Bottom-left: Gesture 3: Redirect light away from hand-gesture. Bottom-right: Gesture 4: System is turned off.

IV. RESULTS AND CONCLUSIONS

In the present *proof-of-concept* implementation six nodes grouped into two clusters were developed to work conjunction with six camera-ready RPi3s integrated across the interior built-environment in order to enable redirection—be it in the form of drawing or deflecting—of direct natural light. The system performed as expected albeit with a number of important limitations. For example, the low-cost cameras used varied in performance depending on the interior lighting conditions as well as with the distance the hand-gesture was effected. Furthermore, the system worked within a rigid controlled environment with known and specified distances between the cameras and the users. Considerable work must be undertaken in order to be able to ascertain robustness when implementing it within a real-life non-controlled environment. Furthermore, although the present setup uses one MCU for every three nodes, future implementations will consider integrating a smaller and cost-effective MCU in each individual node. Redundancy and resilience inform this consideration, as in the present implementation an entire set of nodes ceases to function if its shared MCU malfunctions. Finally, the number of parts of each node is presently being optimized and redesigned in order to minimize the probabilities of mechanical failure.

REFERENCES

- [1] H. Bier, A. Liu Cheng, S. Mostafavi, A. Anton, and S. Bodea, "Robotic Building as Integration of Design-to-Robotic-Production and -Operation," in *Springer Series in Adaptive Environments*, vol. 1, *Robotic Building*, H. Bier, Ed. 1st ed.: Springer International Publishing AG, 2018.
- [2] M. A. Steane, *The Architecture of Light: Recent Approaches to Designing with Direct natural light*. Hoboken: Taylor and Francis, 2012.
- [3] A. Haans, "The natural preference in people's appraisal of light," *Journal of Environmental Psychology*, vol. 39, pp. 51–61, 2014.
- [4] K. C.H.J. Smolders and Y. A.W. de Kort, "Bright light and mental fatigue: Effects on alertness, vitality, performance and physiological arousal," *Journal of Environmental Psychology*, vol. 39, pp. 77–91, 2014.
- [5] K.C.H.J. Smolders, Y.A.W. de Kort, and S. M. van den Berg, "Daytime light exposure and feelings of vitality: Results of a field study during regular weekdays," *Journal of Environmental Psychology*, vol. 36, pp. 270–279, 2013.
- [6] B. K. Hawes, T. T. Brunyé, C. R. Mahoney, J. M. Sullivan, and C. D. Aall, "Effects of four workplace lighting technologies on perception, cognition and affective state," *International Journal of Industrial Ergonomics*, vol. 42, no. 1, pp. 122–128, 2012.
- [7] E. C. Marqueze *et al.*, "Direct natural light exposure, sleep and depression among day workers and shiftworkers at arctic and equatorial latitudes," (eng), *PloS one*, vol. 10, no. 4, e0122078, 2015.
- [8] E. Pasquini, "L'Institut du Monde Arabe-a "Must See" On a Visit to Paris," *The Washington report on Middle East affairs*, vol. 29, no. 2, p. 61, 2010.
- [9] Council on Tall Buildings and Urban Habitat, *Al Bahar Towers—External Automated Shading System: CTBUH Innovation Award Winner*, 2012.
- [10] A. Schlüter, *Soft robotics for adaptive building facades*. Accessed on: Apr. 20 2016.
- [11] A. Liu Cheng and H. Bier, "Adaptive Building-Skin Components as Context-Aware Nodes in an Extended Cyber-Physical Network," in *Proceedings of the 3rd IEEE World Forum on Internet of Things: IEEE*, 2016, pp. 257–262.
- [12] K. Oosterhuis, "Caught in the Act," in *Applied Virtuality Book Series*, v.8, *ALIVE: Advancements in adaptive architecture*, M. Kretzer and L. Hovestadt, Eds, Basel/Berlin/Boston: Birkhäuser, 2014, pp. 114–119.
- [13] A. Liu Cheng, H. Bier, and G. Latorre, "Actuation Confirmation and Negation via Facial-Identity and -Expression Recognition," in *Proceedings of the 3rd IEEE Ecuador Technical Chapters Meeting (ETCM) 2018*, 2018, in press.
- [14] A. Liu Cheng, H. Bier, and S. Mostafavi, "Deep Learning Object-Recognition in a Design-to-Robotic-Production and -Operation Implementation," in *Proceedings of the 2nd IEEE Ecuador Technical Chapters Meeting 2017*, Guayaquil, Ecuador, 2017.
- [15] DT42©, Ltd, *BerryNet®: Deep learning gateway on Raspberry Pi*. [Online] Available: <https://github.com/DT42/BerryNet>. Accessed on: Jun. 21 2017.
- [16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the Inception Architecture for Computer Vision*. Available: <http://arxiv.org/pdf/1512.00567>.
- [17] J. Redmon and A. Farhadi, *YOLO9000: Better, Faster, Stronger*. Available: <http://arxiv.org/pdf/1612.08242>.
- [18] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications," *IEEE Commun. Surv. Tutorials*, vol. 16, no. 4, pp. 1996–2018, 2014.
- [19] OpenCV, *OpenCV: Open Source Computer Vision Library*. [Online] Available: <https://opencv.org/>.
- [20] P. Soffritti, *Hand detection software written in C++ using OpenCV*. [Online] Available: <https://github.com/PierfrancescoSoffritti/Handy>. Accessed on: Jun. 18 2018.