Actuation Confirmation and Negation via Facial-Identity and -Expression Recognition

Alexander Liu Cheng[1,2], Henriette Bier[1,3], Galoget Latorre[4]
[1]Faculty of Architecture and the Built Environment, Delft University of Technology, Delft, The Netherlands
[2]Facultad de Arquitectura e Ingenierías, Universidad Internacional SEK, Quito, Ecuador
[3]Dessau Institute of Architecture, Anhalt University of Applied Sciences, Dessau, Germany
[4]Facultad de Ingeniería de Sistemas, Escuela Politécnica Nacional, Quito, Ecuador

E-mail: {a.liucheng, h.h.bier}@tudelft.nl, galoget.latorre@epn.edu.ec

***Abstract*—This paper presents the implementation of a facial-identity and -expression recognition mechanism that confirms or negates physical and/or computational actuations in an intelligent built-environment. Said mechanism is built via Google Brain's *TensorFlow* (as regards facial identity recognition) and Google Cloud Platform's *Cloud Vision API* (as regards facial gesture recognition); and it is integrated into the ongoing development of an intelligent built-environment framework, viz., *Design-to-Robotic-Production & -Operation* (D2RP&O), conceived at *Delft University of Technology* (TUD). The present work builds on the inherited technological ecosystem and technical functionality of the *Design-to-Robotic-Operation* (D2RO) component of said framework; and its implementation is validated via two scenarios (physical and computational). In the first scenario—and building on an inherited adaptive mechanism—if building-skin components perceive a rise in interior temperature levels, natural ventilation is promoted by increasing degrees of aperture. This measure is presently confirmed or negated by a corresponding facial expression on the part of the user in response to said reaction, which serves as an intuitive override / feedback mechanism to the intelligent building-skin mechanism's decision-making process. In the second scenario—and building on another inherited mechanism—if an accidental fall is detected and the user remains consciously or unconsciously collapsed, a series of automated emergency notifications (e.g., SMS, email, etc.) are sent to family and/or care-takers by particular mechanisms in the intelligent built-environment. The precision of this measure and its execution are presently confirmed by (a) identity detection of the victim, and (b) recognition of a reflexive facial gesture of pain and/or displeasure. The work presented in this paper promotes a considered relationship between the architecture of the built-environment and the *Information and Communication Technologies* (ICTs) embedded and/or deployed.**

***Keywords*—Design-to-Robotic-Production & -Operation, Wireless Sensor and Actuator Networks, Facial Recognition, Ambient Intelligence, Adaptive Architecture**

I. Introduction

This paper is part of a series of discrete—yet incremental—developments that promote *Design-to-Robotic-Production & -Operation* (D2RP&O) [1] strategies as effective alternatives to those of *Ambient Intelligence* (AmI) [2] / *Ambient Assisted Living* (AAL) [3] in enabling intelligent built-environments. AmI's / AAL's approach—as evidenced by a sampling of its literature (e.g., [3–5])—centers around the development of solutions based on *Information and Communication Technologies* (ICTs), where *Architecture, Engineering, and Construction* (AEC) considerations—if entertained—are relegated to the periphery. This results in ICT-based solutions whose sophistication surpasses that of the built-environment's within which they are to be deployed, which often entails high installation costs [6] associated with retrofitting and/or late-stage AEC modifications. Furthermore, according to a recent review, there is no evidence that AmI- / AAL-based *smart home* technologies contribute to health-related quality of life [7].

While D2RP&O subsumes AmI's objective to promote a vision of the future dwelling space as a digital living room empowered by embedded-technologies and capable of promoting user-wellbeing, it differs in its emphasis on early-stage integration of both AEC considerations and ICTs. That is, with respect to AEC, D2RP considers composition, form, optimization, robotic fabrication, and integration of materially heterogenous components partially informed by intended ICT-based mechanisms / services. With respect to ICTs, D2RO considers technical and technological systems pertaining to computational / robotic services to be deployed in the resulting D2RP-informed built-environment. Consequently, decisions adopted in the AEC domain are considered in the computational / robotic and vice versa, resulting in a highly deliberate design strategy where neither form / space nor ICT-based mechanisms / services (physical and computational) are incidental with respect to one another. In this manner, the architecture of a built-environment functions as a fundamental support to its ICT-based mechanisms / services and vice-versa, while the built-environment as a unified whole promotes user-wellbeing.

The present implementation inherits the latest iteration of the system architecture developed by the authors [8], one that considers the built-environment as a highly heterogenous yet cohesive *Cyber-Physical System* (CPS). At the core of this CPS lies a self-healing and meshed *Wireless Sensor and Actuator Network* (WSAN) that correlates sensed or input data (environmental, user-based, or cloud-service provided) with physical and computational adaptations in the built-

978-1-5386-6657-9/18/$31.00

environment—and vice versa—in order to enhance the user(s) quality of life via a variety of mechanisms / services. Two of these mechanisms / services are presently expanded to integrate a facial-identity and -expression recognition mechanism implemented via Google Brain®'s *TensorFlow*™ [9] and Google Cloud Platform®'s *Cloud Vision API* [10], respectively, in order to confirm or to negate actuations effected by their automated decision-making processes.

The first one is a mechanism that drives adaptive and context-aware building-skin components [11] that adapt to interior and exterior environmental conditions as a swarm—that is, where the action / reaction of one affects and is affected by the reaction of all in a manner proportional to proximity and predetermined influence. In this first mechanism, the facial-identity and -expression recognition feature enables it to identify *who* reacts to and *what* [facial expression] is elicited by a particular actuation and to decide accordingly whether to continue or stop with said actuation depending on the user's known preferences and his/her particular facial expression. The second is a mechanism that drives a fall detection and intervention solution [12] that detects collapsed users in a given space and proceeds to notify family and care-takers via automated SMS and email messages. In this second mechanism, the facial-identity and -expression recognition feature enables it to identify *who* falls and with *what* [facial expression] in order to determine the urgency of the event and to react accordingly.

The objective of these expansions is to increase the efficacy of the inherited mechanisms by demonstrating that the facial identification and expression recognition mechanism developed in this paper enhances the decision-making processes of the two inherited mechanisms by enabling them to detect and to factor subtle facial reactions to particular actuations, which ascertain precision, intuitiveness, and appropriateness of the actuations in question.

II. Concept and Approach

As detailed in the *Introduction*, the present implementation inherits a previously developed system architecture [8] whose WSAN enables the two mechanisms presently expanded. While it is not pertinent to describe this system architecture in detail here, in the following subsections a summary of the mechanisms in question is provided, followed by an explanation of the role of facial-identity and -expression recognition in their enhanced functionality. Incidentally, it may be noted that facial identification and recognition mechanisms have been implemented in intelligent built-environments for a variety of purposes (e.g., [13]). But one overarching innovation of the present implementation is that an instance of such mechanisms is used to confirm or negate actuations in the built-environment in an intuitive and nuanced manner, thereby enhancing the way the user(s) interact with their intelligent built-environment. Furthermore, and with respect to technical and technological innovation, the present implementation adopts a modular and distributed approach where the facial identification component is independent of yet interrelated with the facial expression recognition component, with the former being driven by limited and cost-effective local resources and the latter by powerful and proprietary cloud-based resources. That is, the approach attempts to allocate and distribute adequate resources for the demands of each component in a way that the successful performance of one does not interfere with that of the other, even as they are executed in parallel. This is an important point, as in both scenarios of expansion of inherited mechanisms involve both facial recognition components running in tandem in order to yield personalized reactions.

A. Scenario 1: Enhancing the precision and sensitivity of an Adaptive and Context-Aware Building-Skin System

This scenario expands on an adaptive building-skin system that enables an intuitive and responsive interface between interior and exterior spaces with respect to environmental, thermal, acoustic, and user-comfort considerations. Each of its components act as individual, context-aware, sensor-actuator nodes capable of differentiated—yet correlated—actions, reactions, and interactions. Accordingly, as the sensed data of any device is accessible across all devices in a topology of meshed nodes, the computationally processed behavior of any node is potentially informed by and informing of the status of individual and/or sets of other nodes. In this manner, the building-skin is not construed as a mere envelope, but rather as a system comprised of agents that actively and continuously promote user-comfort [11]. In the present implementation, new building-skin nodes are developed and built (see Fig. 1). The mechanism that drives their functionality remains the same, but the new design enables a larger variety of configurations and degrees of aperture / closure than the original design used in the first implementation of the system. In this scenario, embedded sensors within the built-environment feed the WSAN with temperature and humidity data. When these exceed limits prescribed by heating and cooling requirements defined by *Comité Européen de Normalisation* (CEN) Standard EN15251-2007 [14], the building-skin nodes begin to react in a way as to ascertain optimal temperature and humidity data. This reaction is diffused and graduated, both in terms of time of initial execution, duration, and extent of actuation. More explicitly, those nodes closest to identified areas of high-temperature are first to react, engage for a longer period, and actuate to degrees of maximum aperture to enable optimal ventilation. Those nodes farther away react in proportion to their proximity to the high-temperature areas in question, and to their estimated efficacy to support cooling / ventilation efforts by establishing an intake-outturn air-flow pathway. This automated behavior is now influenced by feedback received via the facial-identity and -expression recognition mechanism, a feedback that may serve to confirm or negate the automatically effected actuation. More specifically, when during the moment of reaction to a high-temperature condition the facial-identity and -expression recognition mechanism detects—to a high-degree of probability (>80%)—that sole user A (see Fig. 3, face A) reacts approvingly to the effected actuation, then said actuation is confirmed and proceeds normally.

978-1-5386-6657-9/18/$31.00 ©2018 European Union

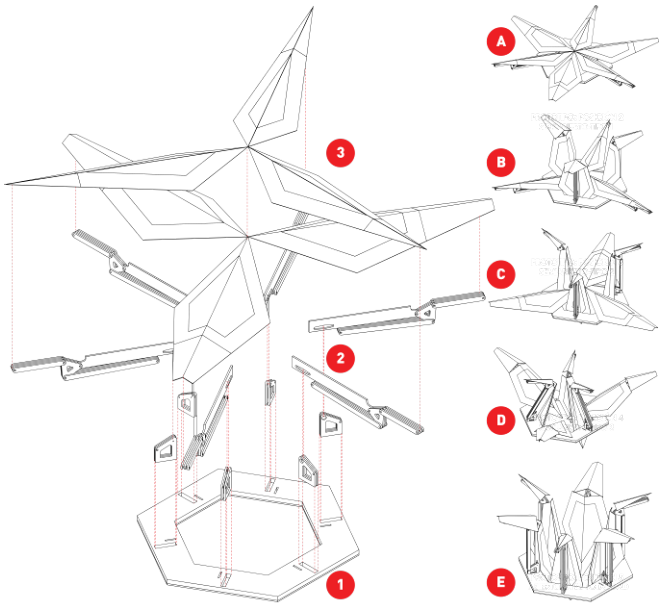

Fig. 1. New adaptive building-skin module concept; items 1: MDF base, with integration of servo motors and a cost-effective Raspberry Pi Camera Module V2; 2: MDF pivoting arms; 3: Opaque acrylic sheets. Possible configurations A: Fully extended, closed; B: Semi-open, variation I; C: Semi-open, variation II; D: Semi-open, hybrid of variations I and II; E: Fully retracted, open.

Alternatively, when said facial recognition mechanism detects that sole user B (see Fig. 3, face B) reacts adversely to the actuation, then it is negated. In situations with more than one user in the built-environment, the decision to confirm or negate a routine actuation comes to a proportional compromise depending on the proximity and reaction of all detected users. That is, suppose the following scenario: Given a region of high-temperature, a set of closest building-skin nodes begins to actuate to maximum degrees of aperture to encourage high-volume ventilation. Users A and B are both in the high-temperature area in question. Accordingly, the facial identification and expression recognition mechanism, via the cost-effective cameras embedded in the building-skin nodes being maximally actuated, begins to ascertain a probable reaction (e.g., joy, sorrow, anger, surprise, etc.) for each user. This probable reaction is given by a straightforward average of the probabilities gathered via each maximally actuating node's embedded camera's data—n.b.: only those probabilities with greater than 80% confidence-level are considered. If user A's reaction is identified as "joy" and user B's reaction as that of "sorrow", then the extent of maximal actuation is mitigated in proportion to the proximity of users A and B to any given building-skin node in maximal actuation. If with respect to a given actuating building-skin node both users are at the same distance, then the actuation will compromise halfway with respect to both—that is, if user A approved of 100% aperture and user B of 0%, then the building-skin node will actuate to 50% aperture. If with respect to another given actuating building-skin node user A is 25% closer than user B, then user A's perceived preference is assigned 25% more weight than that of user B's, and so forth and so on.

B. *Scenario 2: Enhancing the precision and sensitivity of a Fall Detection and Intervention System*

This scenario expands on a fall detection and intervention system developed as a fully operational, real-scale solution [12] that uses two Class 2M 10º line lasers in conjunction with a number of *Light Dependent Resistors* (LDRs) to gauge the probabilities of an emergency-event based on the estimated dimensions of the collapsed object. If the solution construes the probabilities of an emergency event as high, a TurtleBot [15] is sent to the location of the collapsed person and automated notifications are sent to emergency-personnel, care-takers, and/or family via both wireless and cellular technologies. From the previous expansion of this system [16], an object-recognition mechanism is implemented via BerryNet® [17] (built with Inception® ver. 3 [18] for a classification model as well as with TinyYOLO® [19] for a detection model). This object-recognition mechanism uses *Convolutional Neural Networks* (CNNs), which are at the forefront of *Machine Learning* (ML) research [18].

In this implementation, the facial-identity and -expression recognition mechanism (also powered by CNNs, albeit *Multi-Task* CNN—MTCNN) is deployed to complement the functions of the object-detection mechanism. That is to say, while this latter is used to detect kinds of objects (e.g., person, car, cup, book, etc.), the former is used exclusively within the domain of human and probable human-state recognition. In the present expansion, the fall detection and intervention system is integrated with a mechanism capable of detecting *who* the collapsed person is, and *what* facial expression is on his/her face via the same embedded cameras mentioned in the first scenario. Depending on the input of these variables, the urgency and repetition frequency of the system's original intervention mechanisms vary. For example, if the system detects that a known user has fallen and remains collapsed, and that his/her facial expression is construed as "sorrow", "anger", or "surprise" (see Fig. 4), then SMS and email notifications expressing a corresponding urgency-level are sent repeatedly until no object is detected.

III. Methodology and Implementation

The present facial-identity and -expression recognition mechanism is implemented via two independent yet interrelated components. The first—the facial identity recognition component—is implemented locally via Google Brain®'s *TensorFlow*™ [9] (see Fig. 2): while the second—the facial expression recognition component—is implemented via Google Cloud Platform®'s *Cloud Vision API* [10] (see Fig. 4). Additionally, the first component is capable of rudimentary facial expression recognition as well, a feature that is used as a back-up measure in case the second component fails. The second component, however, is incapable of subsuming the function of the first because Cloud Vision API does not support facial identity recognition due to privacy concerns (although it *can* detect human faces). In the implementation of the facial identity recognition component (i.e., the first component), TensorFlow™ is installed on a Linux (Ubuntu) virtual environment and executed in Python.

978-1-5386-6657-9/18/$31.00 ©2018 European Union

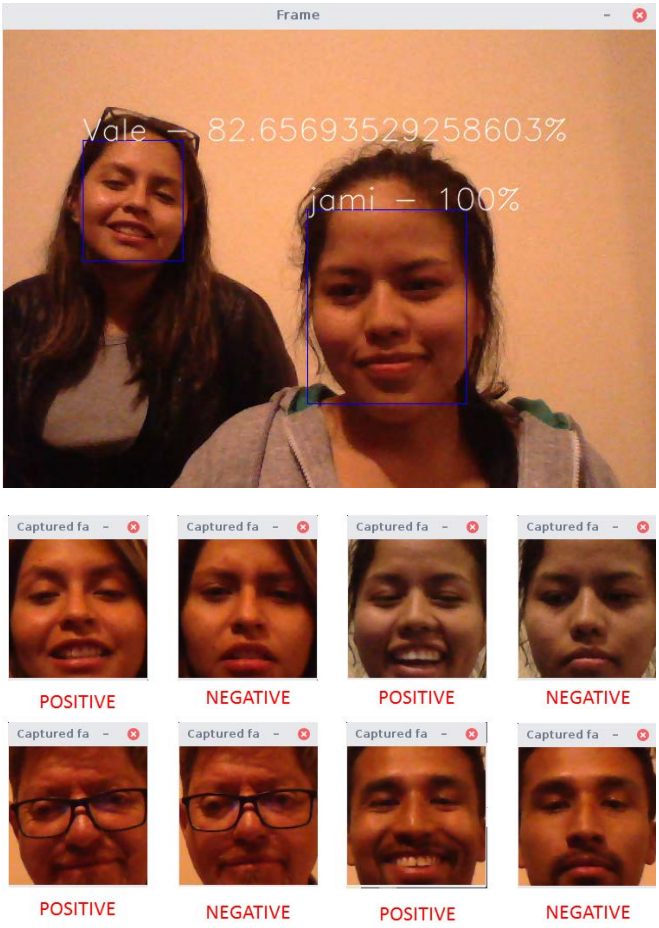

Fig. 2. Top: TensorFlow™ image capture initialization. Middle: real-time Facial Recognition. Bottom: TensorFlow™'s rudimentary "Positive" or "Negative" facial-expression recognition.

During execution of its MTCNN face detection model (see Fig. 2 *Top*), TensorFlow™ requests the user—and, consequently, every individual who is to be subsequently recognized by the system—to let the camera capture his/her face from a variety of positions, orientations, and angles. After completing this phase, facial identity recognition is successfully tested real-time (see Fig. 2 *Middle*). As previously mentioned, TensorFlow™ may also be used for basic facial expression recognition, which serves as a functional back-up in case the component implemented with Cloud Vision API fails. This back-up component is capable of recognizing two broad types of expression: "positive" and "negative" (see Fig. 2 *Bottom*). In the implementation of the facial expression recognition component (i.e., the second component), Python is used to integrate the services of Cloud Vision API into the inherited WSAN. The same visual input is fed to both components to yield a correlated recognition of an identity as well as of a facial expression. This is important, as it is the way that the system knows *who* the recognized facial expression corresponds to (recall that Cloud Vision API does not support facial identity recognition and consequently its returned output by itself is anonymous).

IV. Results and Discussion

A. With respect to Scenario 1's implementation

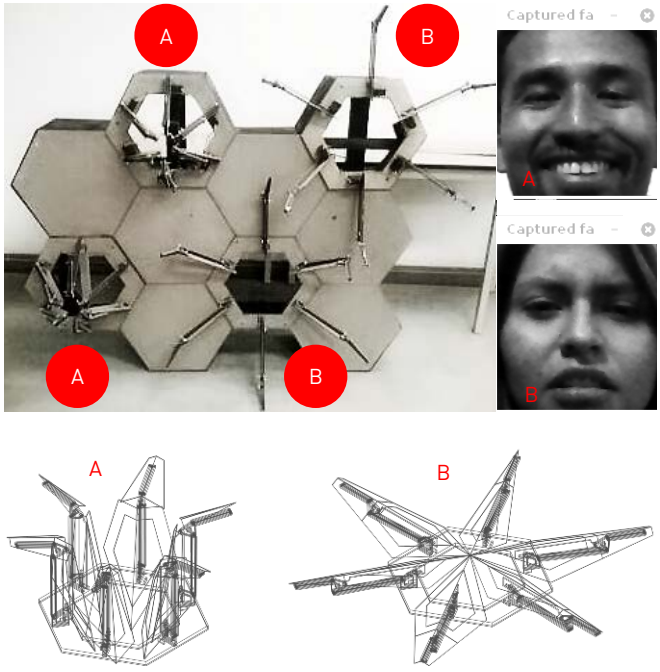

Fig. 3. Top: building-skin fragment reacting differently to agents A and B under tested position configuration 2. Bottom: States of module with respect to agents A (fully open) and B (fully closed).

With respect to the testing and validation of scenario 1, two users—already introduced in previous sections as A and B—with conflicting facial expressions (i.e., user A: joy; B: sorrow), are asked to stand—always aligned to the building-skin fragment's center—(1) next to each other and equidistantly close to the fragment; (2) far from each other and equidistantly close to the fragment; and (3) close to each other and equidistantly far to the fragment, with a simulated high-temperature area around where they stand across all configurations. In the first configuration, all four building-skin modules default to a neutral aperture configuration—i.e., item C in Fig. 1. In the second configuration (see Fig. 3), those modules closest to user A open maximally as the statutory actuation [to open the node] is confirmed by the user's expression; and those modules closet to user B remain shut as the statutory actuation is negated by the user's expression. In the third configuration, due to the large distance between the users and the fragment—and, consequently, between the users and the cameras belonging to the nodes within the fragment—the mechanism is unable to ascertain a high probability favoring a particular facial expression and therefore this consideration is ignored. That is to say, the nodes behave as they would without user feedback. Due to their interesting results, the above three position configurations for users A and B with respect to the fragment are the most salient ones among the other configurations sampled. To some extent, they are also indicative of the most common or anticipated results among the tested configurations. That is to say, more frequently than not—and under the present set-up conditions—the fragment

978-1-5386-6657-9/18/$31.00 ©2018 European Union

defaults to a negotiated average with respect to aperture extent; or to a fair distribution of nodes that satisfy user A and those that satisfy user B; or to a situation where user input / feedback is ignored. These results are more indicative of the experiment's setup—and its limitations—than of the programed behavior of the building-skin module. Nevertheless, such results are sufficient for a modestly successful proof-of-concept implementation.

B. With respect to Scenario 2's implementation

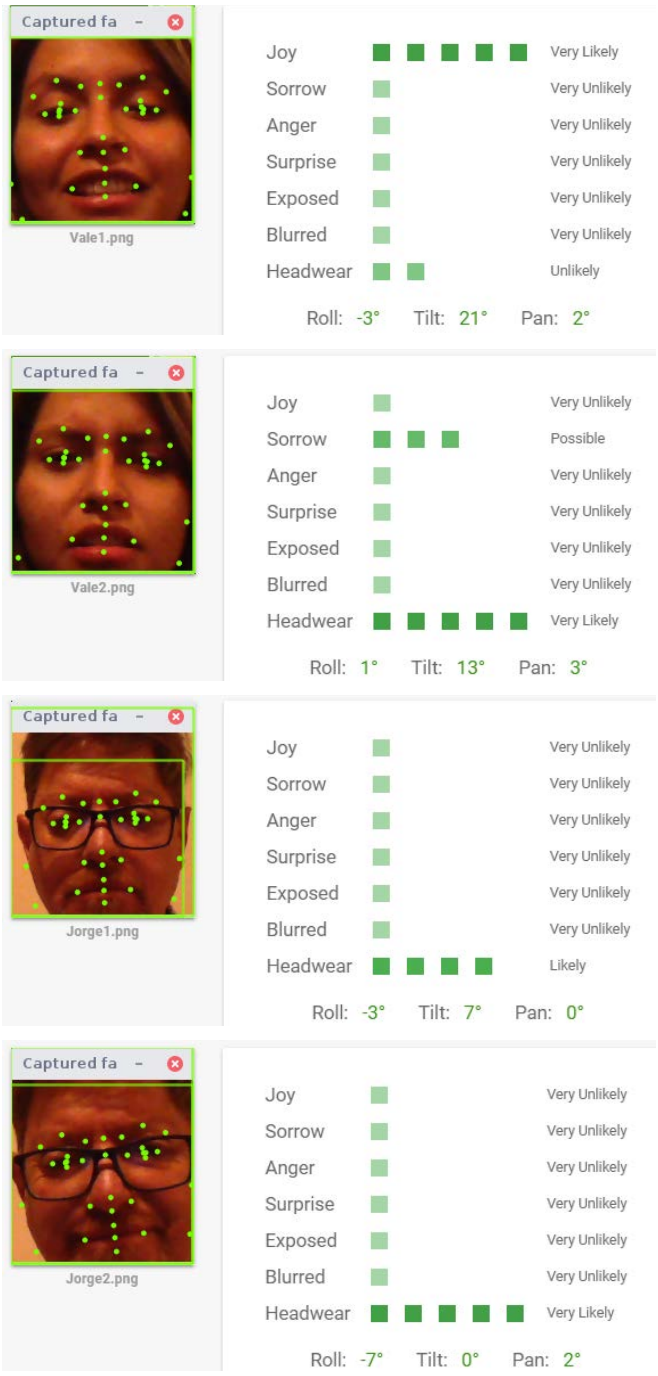

Fig. 4. Top Two: Cloud Vision API's predictions with respect to user B's expressions. Bottom Two: instances of equivocation when glasses are worn.

With respect to the testing and validation of scenario 2, for all instances of successfully detected collapsed people in the original implementation and first expansion of the fall detection and intervention system, a corresponding enacted fall event is produced, where the identity and facial expression of the actor are identified. In this experimental setup, the emphasis of the testing is on the efficacy of the facial-identity and -expression recognition mechanism. In the trial runs, the first mechanism performs successfully by consistently identifying the user and his/her facial expression most of the time (>90%). Instances of false positives occur when hats, glasses, and other accessories are worn (see Fig. 4, *Bottom Two*).

V. Conclusion

This paper presents a facial-identification and -expression recognition mechanism implemented via Google Brain®'s *TensorFlow*™ and Google Cloud Platform®'s *Cloud Vision API*, respectively, in order to confirm or to negate actuations effected by automated decision-making processes in two inherited mechanisms. With respect to the first inherited mechanism, an expansion scenario is developed to consider the identity of users and their facial expressions in reaction to the actuation of building-skin nodes responding to interior and exterior environmental conditions. With respect to the second inherited mechanism, another expansion scenario is developed to consider the identities and facial expressions of the individuals involved in possible accidental fall events. In both scenarios, the facial-identification and -expression recognition mechanism contributes to an increase in precision and nuance with respect to adaptation (in the first scenario) and identification (in the second scenario). Accordingly, said mechanism is construed as a modestly successful addition to the existing ecosystem of mechanisms and services that supervene on the inherited WSAN.

Nevertheless, there are certain challenges and limitations that could be overcome. With respect to challenges, a salient one compels that further work must be carried out to optimize the facial-identification and -expression recognition mechanism with respect to already existing mechanisms and services within said ecosystem in order to avoid unnecessary overlaps in sensors, subsystems, and services (whole or in part). For example, in subsequent implementations, object-recognition activities may be subsumed by services provided by Cloud Vision API, as opposed to the present setup where object-recognition via BerryNet® is placed to work in tandem with facial expression recognition via Cloud Vision API. This setup is informed more by economic considerations than concerns for efficiency and efficacy. BerryNet® is a free object-recognition mechanism that, although not as powerful as other paid alternatives, is nevertheless effective. Notwithstanding this consideration, Cloud Vision API is also free within a certain number of feature-usages, and affordable beyond this limit [10]. More tests are required to ascertain whether efficiency and efficacy may be enhanced by relegating object-recognition tasks to it.

978-1-5386-6657-9/18/$31.00 ©2018 European Union

ACKNOWLEDGMENT

This paper has profited from the contribution of TUD Robotic Building researchers, tutors, and students. Additionally, the authors specifically acknowledge Francisco Cevallos and Jorge Jarrín (of Estudio 685), Valeria Monar and Jamilet Galarza for their assistance in the implementation of the facial identification and expression recognition mechanism; and Soledad Alvares, Gabriela Herrera, Esthefania Quito, and Héctor Solís for their assistance in the implementation of the functional building-skin fragment.

REFERENCES

[1] H. H. Bier, "Robotic Building as Integration of Design-to-Robotic-Production & Operation," *Next Generation Building*, no. 3, 2016.

[2] E. Zelkha, B. Epstein, S. Birrell, and C. Dodsworth, "From Devices to 'Ambient Intelligence': The Transformation of Consumer Electronics (Conference Keynote)," in *Digital Living Room Conference*, 1998.

[3] R. Wichert and H. Klausing, Eds., *Ambient Assisted Living 8. AAL-Kongress 2015 Frankfurt/Main, Germany, April 29-30, 2015*. Berlin/Heidelberg, Germany: Springer Berlin Heidelberg, 2015.

[4] P. Novais, K. Hallenborg, D. I. Tapia, and J. M. C. Rodríguez, *Ambient Intelligence - Software and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[5] H. Lindgren *et al., Ambient intelligence: Software and applications -- 7th International Symposium on Ambient Intelligence (ISAmI 2016)*. Switzerland: Springer, 2016.

[6] R. Wichert, F. Furfari, A. Kung, and M. R. Tazari, "How to Overcome the Market Entrance Barrier and Achieve the Market Breakthrough in AAL," in *Ambient Assisted Living 5. AAL-Kongress 2012 Berlin, Germany, January 24-25, 2012*, R. Wichert and B. Eberhardt, Eds., Heidelberg, New York: Springer, 2012, pp. 349–358.

[7] L. Liu, E. Stroulia, I. Nikolaidis, A. Miguel-Cruz, and A. Rios Rincon, "Smart homes and home health monitoring technologies for older adults: A systematic review," (eng), *International journal of medical informatics*, vol. 91, pp. 44–59, 2016.

[8] A. Liu Cheng and H. H. Bier, "Extension of a High-Resolution Intelligence Implementation via Design-to-Robotic-Production and -Operation strategies," in *Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC 2018)*, Berlin, Germany, 2018.

[9] TensorFlow™, *An open source machine learning framework for everyone.* [Online] Available: https://www.tensorflow.org/. Accessed on: Apr. 20 2018.

[10] Google Cloud Platform®, *Cloud Vision API: Derive insight from images with our powerful Cloud Vision API.* [Online] Available: https://cloud.google.com/vision/. Accessed on: Apr. 20 2018.

[11] A. Liu Cheng and H. 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] A. Liu Cheng, C. Georgoulas, and T. Bock, "Fall Detection and Intervention based on Wireless Sensor Network Technologies," *Automation in Construction*, 2016.

[13] L. Y. Mano *et al.,* "Exploiting IoT technologies for enhancing Health Smart Homes through patient identification and emotion recognition," *Comput.Commun.*, vol. 89-90, pp. 178–190, 2016.

[14] Comité Européen de Normalisation© (CEN), *Standard EN 15251–2007: Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics.* Accessed on: 16/07/07.

[15] C. Georgoulas, T. Linner, A. Kasatkin, and T. Bock, "An AmI Environment Implementation: Embedding TurtleBot into a novel Robotic Service Wall," in *Proceedings of the 7th German Conference on Robotics*, Munich, Germany: VDE Verlag, 2012.

[16] A. Liu Cheng, H. 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.

[17] DT42©, Ltd., *BerryNet®: Deep learning gateway on Raspberry Pi.* [Online] Available: https://github.com/DT42/BerryNet. Accessed on: Jun. 21 2017.

[18] 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.

[19] J. Redmon and A. Farhadi, *YOLO9000: Better, Faster, Stronger.* Available: http://arxiv.org/pdf/1612.08242.