Augmented Reality Application Supporting On-Site Secondary Building Assets Management

Available online at 2019.creative-construction-conference.com/proceedings/

CCC 2019

Proceedings of the Creative Construction Conference (2019) 110

Edited by: Miroslaw J. Skibniewski & Miklos Hajdu

https://doi.org/10.3311/CCC2019-110

Creative Construction Conference 2019, CCC 2019, 29 June - 2 July 2019, Budapest, Hungary

Augmented reality application supporting on-site secondary building assets management

Berardo Naticchia[a], Alessandra Corneli[a,*], Alessandro Carbonari[a], Frédéric Bosché[b], Lorenzo Principi[a]

[a]*Università Politecnica delle Marche, DICEA, Ancona,60131, Italy*
[b]*Heriot-Watt University, Edinburgh, UK*

Abstract

Secondary building assets management requires a large amount of information related to them. Nevertheless, building assets surveys are cost and time demanding, especially because they need long post-processing efforts in order to systematize collected data. Furthermore, with the recent transition towards the BIM methodology for building management also modeling building objects both in their geometric features and in their related information is a long process and error-prone task. Under these circumstances the possibility of performing the majority of operation on-site would definitely make the process more efficient and it would reduce errors. Augmented Reality (AR) with its capability of overlapping digital data to the real scene is the right tool to support operators on-site.

The proposed system has the aim of reducing the time of secondary building assets survey and provide a tool for the automatic enrichment of BIM models. An AR device (Hololens) with an embedded computer and a neural compute stick constitute the portable on-site system for the automatic recognition of assets objects, removing the necessity of reworking data off site. A trained Deep Learning Neural Network inside the neural compute stick performs the recognition providing the operator with objects features and position. The AR application inside the Hololens operates as an interface between the user and the digital information just created. Finally, data is stored in a NoSQL database linked to the BIM model so as to be available for further operations. The visually supporting information provided by the AR tool, the possibility of working on data directly on site and the portability of the system represent means for increasing efficiency in survey operations. First tests have been conducted so as to prove the feasibility of the system and its use on site without further equipment.

Keywords: Augmented Reality; BIM; Facility Management; Building asset.

MR	Mixed Reality	AI	Artificial Intelligence
AR	Augmented Reality	AIM	Asset Information Model
NN	Neural Network	FM	Facility Management
BIM	Building Information Model		

*Corresponding author: a.corneli@pm.univpm.it

1. Introduction

Information management represents a big challenge in the AEC industry especially during buildings operational phase [1, 2]. The cost of information is doubling since it is paid once during the project and after during the building lifecycle to retrieve all the necessary data, which can also be not available, mainly in existing constructions [1, 3, 4]. A number of techniques and technologies are now in use, including EDM (Electronic Distance Measurement), GPS (Global Positioning System), 3D Laser scanner [5]. However, each of the aforementioned technologies requires further post processing of data in order to provide interpretation. On the other hand, BIM is spreading as a standard for information management. For this reason, even Facility Management (FM) is expected to be based on an Asset Information Model (AIM) populated from the BIM model which would play a beneficial role in many FM practices [6]. Anyway, data is still often collected manually while the digitization that the construction industry is facing in recent years has led to a growing interest in one of the major benefits of this change: the automation of processes. For the reasons stated above with this research we propose a system that exploits the advantages of different technologies such as Mixed Reality, BIM and Neural Network with the aim of reducing post-processing effort in the interpretation of data thanks to the automation of some processes and an efficient human-machine collaboration.

2. Scientific Background

Building surveys still demand high costs and time to be pursued since, despite the high accuracy achieved by the latest techniques, such as laser scanning and photogrammetry, much of the work is still done post processing. Among the latest procedures proposed Adan et al. presented a method for the detection of ‘small components’ such as fire extinguishers, switches, sockets and signs [7] through 3D laser scanned point clouds. On the other hand, Lu et al. [8] recognize and model building structural components with a combination of three sub-systems: 1) object recognition, based on a new neuro-fuzzy framework; (2) material recognition, based on image classification procedures; (3) IFC BIM object generation. All of the aforementioned methods still require a post processing effort and data interpretation pursued off site without the possibility of checking or adding missing information. Bonanni et al. [9] proposed a method that works on site combining a robot, which employs a SLAM (Simultaneous Localization and Mapping) module for building up the map of the environment, and human input to provide spatial hints about entities of interest. This solution is still laborious since the operator has to manually detect the object and enter its features. On the other hand, MR has already demonstrated its potentiality in supporting operations on buildings [10] displaying information through holograms which is a much more convenient way to bring data on site [11]. One example of this is represented by Ammari et al. [12] who developed a system capable of showing holograms thanks to Augmented Reality (AR) and image tracking. Kopsida and Brilakis [13] present an evaluation of different methods that could be implemented for a marker-less mobile BIM-based AR solution for inspections, concluding that there are no efficient mobile AR solutions for on-site inspections and that other methods for marker-less AR, even if already introduced, have not yet been tested on construction sites.

Lastly in recent years there have been already done tests about the combination use of Neural Network (NN) and Augmented Reality together. Baek et al. [14] proposed a NN-based method for indoor localization with the purpose of providing relevant information in FM applications. Nevertheless, many issues have still to be addressed: the proper scale of the visualization of building components in situ; accurate localization of the operator inside the building so as to automatically display relevant information; recognizing building assets without specific markers (visual markers or RFID tags); and providing an effective interface between MR headsets and BIM data.

3. Secondary building assets survey

The system proposed has been specifically thought for the use case of secondary building assets survey. The inventory is a costly operation [2] since it requires a large amount of money and thousands of man-hours for creating/updating information [3, 4] necessary for operating buildings. Current inventory procedures still require long post-processing [15], while the system proposed aims to reach a certain level of automation in the process of data collection improving the efficiency and leading to an automatic enrichment of BIM models.

The system presented with this research proposes a new survey procedure. The operator wears the MR head-mounted display and walks into the building for the first time. The recognition application developed is able to recognize secondary components just framing the objects. The snapshots of the Hololens is sent in real time to an embedded system where NN carry out the recognition process whose result is a label associated with a bounding box that contains the recognized object. The label is the piece of information that the application uses to choose a specific object among a predefined library. The bounding box, which is defined by the x,y coordinates of its center plus width and height, is used to place the object, as a hologram, into the real environment. When the application is not able to recognize the object there is still the possibility of adding it manually in the correct position. At the end of the survey all the data and positions of the detected objects are transferred into a NoSQL Database so as to be available for future operations.
One of the most valuable advantages of this procedure lays in the checking real time the collected data with the possibility of modifying them if necessary directly on site. This will reduce the possibility of incorrect or missing data since all the survey operations are accomplished on site, avoiding a post interpretation that is an error prone procedure. On the other hand, the AR feature of displaying the gathered information overlapped to the real world make it easier to interact with them. In contrast with current methods where there is a prevalence of machine-human working in series the proposed method will lead to a machine-human work in parallel system reaching a higher efficiency.

4. Recognition system components

The system components can be grouped in two different environments: the MR environment which is where the digital copies of the real objects are developed including the object recognition application explained in the following paragraph and the real environment which is represented by the building and the assets object of the survey, the operator and his whole equipment. The MR environment is composed of the following components (Figure 1):

- BIM environment used to develop the initial building model.
- The MR platform (Unity software) that allows to develop applications for the MR tool.

The real environment, on the other hand, comprises the following elements (Figure 1):

- Microsoft Hololens which is the MR head-mounted display. It acts as an interface between the digital world and the real world (operator and environment).
- On-site operator who is doing the survey wearing the Hololens and the embedded system.
- Neural Computer Stick-Movidius which is specially designed for working with neural networks. In order to make the entire system usable on site the images are processed by this tool, as opposed to rely on cloud computation.
- Embedded PC- Raspberry which works as an interface between the Hololens and the Movidius, and as a hardware support for the latter.

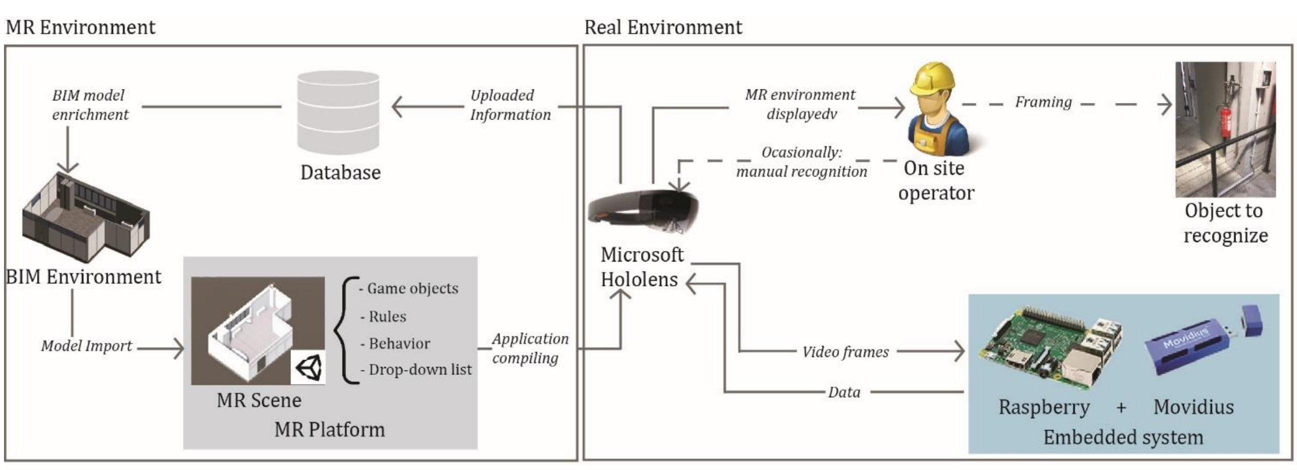

Figure 1 - System components

4.1. Recognition application

The recognition application will be developed in Unity, using the programming language C#. This application will carry out the following tasks:

1.to read and interpret the data about the position of the operator inside the building, and therefore in its digital twin;
2.to send the snapshots to the embedded system;
3.to read the data from the recognition process (bounding box coordinates and object type);
4.to identify among a predefined library of objects the object type that matches the recognition response;
5.to locate the object in the right position according to the bounding box coordinates provided in the recognition response and the depth dimension provided by the mesh (that Hololens automatically does because the fire extinguisher owns a behavior for its positioning on walls);
6.to provide the possibility of modifying the object type or its position manually;
7.to provide the possibility of adding objects manually.

Tests on the insertion of doors and windows in the MR environment have been carried out.

4.2. Neural network

Neural Networks represent the method chosen for the recognition of small objects. In order to perform the recognition with NN it is just necessary to frame the object; this simplicity in their usage make their use on site fast and instinctive. For our purposes the NN not only has to recognize the correct type of object but also its position within the frame, so as to make it possible its positioning in the real environment. Among all the types of neural networks that exist the YOLO, which are able to perform classification and localization in one-step, is the one chosen for this project. This choice depends upon the speed of this kind of NN which is 45 frames per second; making the snapshots processed in real-time, with negligible latency of a few milliseconds. Furthermore, the YOLO can predict multiple bounding boxes and scores simultaneously. Finally, they are an open source solution [16, 17]. In this project, a pre-trained YOLO is used. In order to customize the recognition process it is possible to re-train the last level of the network. The creation of the necessary dataset of images to train the network for our purpose is explained in the following sub-section.

4.2.1. Fire extinguisher datasets

The dataset to train the network should have specific features and it has to include at least one image for every existing type of object (fire extinguisher in this case), so as to be able to recognize it no matter the external appearance. The more the number of images the better the network will train. For this reason, the dataset will be made up of both original pictures and graphically re-edited photos as suggested by studies on dataset creation [18]. The creation of the dataset involves also labelling all the images and drawing the bounding box around the object to be recognized. For this operation we worked using Visual Object Tagging Tool. The output of this process is a .txt file for each image with a line declaring the class and the bounding box coordinates (X,Y of the center and width and length). Our first tests aim to recognize only the object, but next steps will handle the recognition of different fire extinguisher type. It is likely that for this purpose the distinctive components of the fire extinguisher (pressure gauge, horn) need to be recognized separately to help in defying the type.

4.2.2. Neural network training

A tiny YOLOv2 pretrained with the COCO dataset was chosen for the first training session. To train the network the following files are required [18, 19]: a file containing all classes labels (1 for our case); a text file with the path to all the images basis of the training; a text file containing all the paths to the previous files; a configuration file with all the layers of YOLO architecture and finally pre-trained convolutional weights. The original *.cfg* file has been modified as follows for our customized training: 1) batch=64, this means we will be using 64 images for every training step; 2) subdivision=8, the batch will be divided by 8; 3)classes=1, the number of categories we want to detect; 4) filters=30, this come from this formula filters=(5+number of classes)×number of bounding box in each grid cell of the image; 5) learning rate=0.001, advised by the developer of YOLO in order to avoid false minimum point. After the training session, the new *.cfg* and weights files are created.

4.2.3. Neural Network testing

First tests about the training of the Neural Network have been carried out on a computer in this preliminary phase. The YOLO tiny v2 chosen has been trained with three different datasets as follows:

The three datasets are composed as follows:

• DATASET 1 (D1) =300 images, 75 (25%) original taken inside the Engineering Faculty premises (Polytechnic University of Marche), 225 (75%) obtained through the augmentation process;

• D1 added to DATASET 2 (D2) = 200 images, 50 (25%) original downloaded from Flickr (only images of fire extinguisher meeting the requirements expressed in the paragraph 5.4.1), 150 (75%) re-edited images;

• D1 and D2 added to DATASET 3 (D3) = 200 images, 50 (25%) original taken inside the Economic and Science Faculties (Polytechnic University of Marche), 150 (75%) re-edited images.

For every dataset the images were divided into two groups, one for the training (75%) and one for the test (15%). The kind of chosen images was similar for all the photos, close-up and with the object entire and placed in the center (Figure 4). The reached mean average precision (mAP) was 60,23% in the first case. The second training counts 86700 iterations and in this case the mAP was 61,54%. In the final training the number of iterations reached 26900 with a mAP of 62,81%. The validation tests have been conducted for each of the three training sessions with the third training dataset. The percentage of fire extinguisher identified started from 38% and reached 45% with the latest training. the precision improved too because while in the first case 26 false positives occurred, in the latest case the number decreased to 14.

4.3. Embedded system

The embedded system is composed by the Neural Compute Stick Movidius and a Raspberry and its aim is to have all the necessary components for performing the survey directly on site. This will make the system self-sufficient and portable in every working situation.

4.3.1. Data transfer from Hololens to computer

Each frame captured from Hololens camera video stream is sent to Raspberry. Single frames are preferred to video stream for the following reasons: i) it's easier to associate each photo with its 3D-information obtained at the instant of the shot; ii) no need for the support of Real Time Protocol required by video streaming; iii) neural networks input for object recognition are single frames. Each frame captured by the Hololens goes through the following steps: i) sending pre-processing; ii) sending; iii) receiving; iv) sending post-processing; v) pre-processing for neural networks; vi) neural network inference; vii) send back NN output to Hololens. Steps (i) and (iv) could be jpeg-compression, cropping and/or scaling. Steps (v) depends on the NN. Exchanging data between Hololens and Raspberry required the development of a custom socket over a Wifi network between Raspberry (hotspot) and Hololens (client). The socket could be either TCP or UDP. Table 1 shows possible configurations. Data rate is the worst, 30 MBps, its measured range is about 50 ± 20 Mbps (uncertainty is due to unmanageable environment variables in both TCP or UDP socket).

Table 1 - Performance tests, each one elaborate 150 photo. (a) Total time for sending raw photos (cropping and scaling times are neglected, max 10 ms). (b) cropped from 896x504 photo. (c) scaled from 896x504.

	Size [kb]		Times [ms]			
Resolution	raw	jpeg	encoding	sending	total	total[(a)]
1408x792	3'340	154	94	5	99	111
896x504	1'350	60	46	2	91	45
416x416[(b)]	519(b)	24	29	0,8	46	17
416x416[(c)]	519	32	45	1,1	62	17
416x234 [(c)]	292(b)	23	36	0,8	46	10

4.3.2. Object detection result transfer

The chosen NN YOLOv2 unit is a float number, received frames are in byte format, a pre-processing step, *(v)*, represents the conversion from byte to float. The YOLOv2 input is a 416x416x3 tensor, corresponding to the frames to be processed (416x416 is resolution, 3 is number of float for pixel). The network (loaded on Movidius) processes this input and returns an output of 13x13x[(*C*+5)x5]. The frame is divided in 507 cells of same width and height, for each cell it finds 5 bounded boxes and for each of them it calculates: width, height, horizontal and vertical distance from top left corner; objectiness which is the probability of an object to rely in this bounding box (not a particular type of object, just an object); *C* probabilities indicating the object category in the bounding box. Then it filters bounding boxes with objectiness lower than a fixed threshold and, for the rest ones, filter those object categories having probability lower than an arbitrary threshold. If the purpose is to recognize just fire-extinguisher, the Raspberry sent back to Hololens one bounding box (through the socket) defined by x,y,w,h and the photo id (used by Hololens to retrieve the saved 3d information of the frame to place the hologram in the three-dimensional space). Filtering will take place only in Raspberry to reduce Hololens work. Due to the small size (could be 20 bytes) of Rasp output, sending time could be neglected.

5. Conclusion and further steps

This research aims to provide a support to FM operations during the building lifecycle in order to improve efficiency. Three different technologies are combined together for an on-site use: the AR, the BIM data model and Deep Learning. First feasibility tests have been conducted related to the NN and data transfer. Next steps will include the improvement of NN performances, the recognition of precise type of fire extinguisher and the full development of the AR interface and tests in a real environment.

References

[1] East E.W., Brodt W., 2007, BIM for Construction Handover, Journal of Building Information Modeling.
[2] Lee S., An H., Yu J., 2012, An extension of the technology acceptance model for BIM-based FM, Construction Research Congress, 602-611. https://doi.org/10.1061/9780784412329.061
[3] Keady R. A., 2013, Financial Impact and analysis of equipment inventories, Facilities Engineering Journal.
[4] Gallaher M.P., O′ Connor A.C., Dettbarn J.L. & Gilday L.T., 2004, Cost analysis of inadequate interoperability in the U.S. capital facilities industry, U.S. Department of Commerce Technology Administration, (NIST GCR 04-867). https://doi.org/10.6028/NIST.GCR.04-867
[5] Arayici Y., 2010, Use of the 3D laser scanner technology in the Built Environment, in Geller, P.S. (Ed.), Built Environment: Design, Management, and Applications, Environmental Science, Engineering and Technology, Nova Science Publishers, New York, NY, 89-120.
[6] Burcin Becerik-Gerber, Farrokh Jazizadeh, Nan Li, and Gulben Calis, 2012, Application Areas and Data Requirements for BIM-Enabled Facilities Management, Journal of Construction Engineering and Management, Vol. 138, Issue 3. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000433
[7] Quintana B., Prieto S. A., Adán A. & Bosché F., 2017, Scan-to-BIM for small building components, Proceedings of the Joint Conference on Computing in Construction (JC3), Heraklion, Greece. https://doi.org/10.24928/jc3-2017/0139
[8] Lu Q, Lee S, Chen L., 2018, Image-driven fuzzy-based system to construct as-is IFC BIM objects. Automation in Construction.; 92:68-87. https://doi.org/10.1016/j.autcon.2018.03.03
[9] Bonanni, T. M. et al., 2013, Human-Robot collaboration for semantic labeling of the environment, Cognitive Cooperating Robots, pp. 1-25.
[10] Riexinger, Günther, et al., 2018, Mixed Reality for On-Site Self-Instruction and Self-Inspection with Building Information Models. Procedia CIRP, 72: 1124-1129. https://doi.org/10.1016/j.procir.2018.03.160
[11] Fonnet, A., Alves, N., Sousa, N., Guevara, M.A., Magalhães, L., 2017, Heritage BIM integration with mixed reality for building preventive maintenance, 24º Encontro Português de Computação Gráfica e Interação (EPCGI), 1-7. https://doi.org/10.1109/EPCGI.2017.8124304
[12] AMMARI, Khaled El; HAMMAD, Amin, 2014, Collaborative BIM-based markerless mixed reality framework for facilities maintenance. In: Computing in Civil and Building Engineering, p. 657-664. https://doi.org/10.1061/9780784413616.082
[13] Kopsida, M., Brilakis, I., 2017, BIM Registration Methods for Mobile Augmented Reality-Based Inspection. In: eWork and eBusiness in Architecture, Engineering and Construction: ECPPM 2016: Proceedings of the 11th European Conference on Product and Process Modelling (ECPPM 2016), Limassol, Cyprus, 7-9 September 2016. CRC Press, p. 201. https://doi.org/10.17863/CAM.42274
[14] Baek, F., Ha, I., & Kim, H. (2019). Augmented reality system for facility management using image-based indoor localization. Automation in Construction, 99(August 2018), 18–26. https://doi.org/10.1016/j.autcon.2018.11.034
[15] Adán, A., Quintana, B., Prieto, S. A. & Bosché, F. (2018). Automation in Construction Door detection in 3D coloured point clouds of indoor environments. Automation in Construction, 85(October 2016), 146–166. https://doi.org/10.1016/j.autcon.2017.10.016
[16] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 779-788. https://doi.org/10.1109/CVPR.2016.91
[17] Shinde, S., Kothari, A., Gupta, 2018, YOLO based Human Action Recognition and Localization. Procedia Computer Science, 133, 831–838. https://doi.org/10.1016/j.procs.2018.07.112
[18] Mas L. Q., Montserrat D., Allebach J., Delp E. J., 2017, Training Object Detection and Recognition CNN Models Using Data Augmentation, Electronic Imaging, Imaging and Multimedia Analytics in a Web and Mobile World 2017, pp. 27-36(10).