Augmented Reality and Deep Learning towards the Management of Secondary Building Assets

A. Corneli[a]*, B. Naticchia[a], A. Carbonari[a], and F. Bosché[b]

[a]Polytechnic University of Marche, DICEA, Ancona, Italy
[b] Heriot-Watt University, Edinburgh, UK
E-mail: a.corneli@pm.univpm.it; b.naticchia@univpm.it; alessandro.carbonari@staff.univpm.it; f.n.bosche@hw.ac.uk

Abstract -

The retrieval of as-is information for existing buildings is a prerequisite for effectively operating facilities, through the creation or updating of Building/Asset Information Models (BIM/AIM), or Digital Twins. At present, many studies focus on the capture of geometry for the modelling of primary components, overlooking the fact that many recurring actions need to be conducted on assets inside buildings. Furthermore, highly accurate survey techniques like laser scanning need long offsite processing for object recognition. Performing such process on site would dramatically impact efficiency and also prevent the need to revisit the site in the case of insufficient/incomplete data.

In this paper, an Augmented Reality (AR) system is proposed enabling inventory, information retrieval and information update directly on-site. It would reduce post-processing work and avoid loss of information and unreliability of data. The system has a Head-Mounted Display (HMD) AR interface that lets the technician interact hands-free with the real world and digital information contained in the BIM/AIM. A trained Deep Learning Neural Network operates the automatic recognition of objects in the field of view of the user and their placement into the digital BIM. In this paper, two uses cases are described: one is the inventory of small assets inside buildings to populate a BIM/AIM, and the second is the retrieval of relevant information from the AIM to support maintenance operations. Partial development and feasibility tests of the first use case applied to fire extinguishers, have been carried out to assess the feasibility and value of this system.

Keywords -

Augmented Reality, Building survey, Inventory, BIM, Facility Management.

1 Introduction

One of the biggest challenges in the construction industry today is information management both in the construction phase and in the operational phase [1, 2].

In fact, the retrieval of specific data during the lifecycle of buildings represents a high cost for all the stakeholders involved in this field [3, 4]. Today, facility maintenance contractors are paid to survey existing buildings in order to capture the as-built/as-is status. In these cases, owners would pay twice, once for the construction contractor to complete the documents at the end of construction and again for the maintenance contractor survey [1]. Moreover, as far as existing buildings are concerned, information is often not available nor updated, although data mirroring the current state (i.e. digital twin) are needed for further action. As a result, a building survey is usually necessary. 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 its interpretation. This may also require further site visits to acquire supplementary data.

Another aspect to take into consideration is how the information is stored after being collected. Since BIM is becoming an industry standard, Facility Management (FM) is expected to be based on information contained in an Asset Information model (AIM) populated from the BIM model. Information collected through a BIM process and stored in a BIM/AIM could be beneficial for a variety of FM practices: such as commissioning and closeout, quality control and assurance, energy management, maintenance and repair, and space management [6]. Furthermore, data modelling is often conducted manually. The automatic detection of objects types and the related automatic creation of the digital objects would improve process efficiency and data reliability.

For the reasons stated above the objectives of this research are as follows:

- to decrease post-processing efforts thanks to the automation of some processes and smart human intervention on site;
- to reduce the time necessary to gather any piece of information;
- to collect data in in a BIM model by creating standard BIM objects with linked data including all information necessary for further operations.

The core innovation proposed in this research lays in the integration of several hardware and software innovations into one system architecture: Neural Networks, Mixed Reality (MR) and BIM. The proposed system develops efficient human-machine collaboration, employing MR as a powerful medium between the human, (cloud) computing facilities and the BIM data.

2 Literature review

In recent years there have been increasing efforts towards the automation of survey procedures. Technologies such as Laser Scanning and photogrammetry are very accurate but they also collect a vast amount of raw data that need to be interpreted. Solutions for semi-automatic survey of the environment have already been pursued:

- The Bonanni et al. [7] combine 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 that must be included into the map. Solutions of this type still demand a lot of effort and long time to detect all the objects since the operator has to manually detect the object and enter its features.
- Adan et al. [8] resent a coloured 3D laser scanned point clouds algorithm that allows the location of doors. This system integrates the analysis of both geometry and colour provided by a calibrated set of 3D laser scanner and colour camera with the aim of detecting, localizing and sizing doors.
- The method proposed by Quintana et al. [9] regards a system for the detection of 'small components' in coloured point clouds acquired by a 3D laser scanner.
- Lu et al. [10] propose a recognition system composed of sub-systems for: (1) object recognition, based on a new neuro-fuzzy framework; (2) material recognition, based on image classification procedures and the trained texture library; (3) IFC BIM object generation, that automatically transforms recognized objects with materials information into complete BIM objects in IFC.

These studies focus on the automatic interpretation of collected data. Anyway, they still require post processing effort, such as the Adan [8] and the Quintana [9] studies, or they are way too complex to be effective during on site procedures, like the Bonanni’s one.

As far as the support given by MR in maintenance operations is concerned some early experiments have already demonstrated the possibilities provided by the display of information on site. The numerous advantages in the use of MR towards the entire lifecycle of building is well described by Riexinger et al. [12]. Both this article and the one by Fonnet et al. [13] demonstrate the convenience in displaying information on site through holograms so as to help during the refurbishment of buildings. Practically, Ammari et al. [11] developed a system capable of showing holograms thanks to Augmented Reality (AR) and image tracking. The system is composed of a IFC XML database that makes it possible to display separated pieces of information if necessary. Hugo et al. [14] used Revit for the geometric and appearance features and Dynamo for a query of data transferred to Unity in XML format. Then they developed the MR environment in Unity and finally deployed into the Hololens a first attempt of holographic projection of a building three dimensions model at a small scale. Finally, Kopsida and Brilakis [15] present an evaluation of different methods that could be implemented for a marker-less mobile BIM -based AR solution for inspections. In conclusion they state 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.

As for the use of Neural Network (NN) and Augmented Reality together, Baek et al. [16] proposed a NN-based method for indoor localization. Their work is also motivated by the need to provide relevant information in FM applications.

While these works demonstrate the growing interest of the AEC industry in MR, 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.

In this paper we propose an architecture that combines some of the aforementioned techniques to allow operators effectively survey buildings for maintenance applications, with MR interface to the

BIM/AIM model, and automatic object detection supported by NN.

A significant challenge to the delivery of such system is the development of an effective approach to indoor user localization.

3 System Architecture

The system architecture supports both use cases previously defined and it can be divided into two main environments as shown in Figure 1:

- The Mixed Reality (MR) environment which is where the digital copies of the real objects are developed and/or manipulated while on siteThis includes an object recognition application (paragraph 5.2).
- The real environment which is represented by the building and the assets object of the survey, the operator and its whole equipment. The real environment is the place where the holograms are displayed and overlapped to it.

The MR environment is composed of the following components:

- BIM environment used to develop the initial building model. This digital twin of the real building will be enriched with the data from the survey.
- The MR environment, which is represented by the software Unity, that allows to develop applications for the MR tool. The scene inside Unity includes the building model imported and all the components necessary to make operations and have interactions between reality and digital information.
- The real environment on the other hand comprises the following elements:
- Microsoft Hololens which is the head-mounted display chosen to show the MR environment on site and to act 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.
- Neural Computer Stick-Movidius which is a neural computer stick that 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.

The Hololens application created in Unity performs the following tasks:

1. it keeps track of the position of the gaze (the

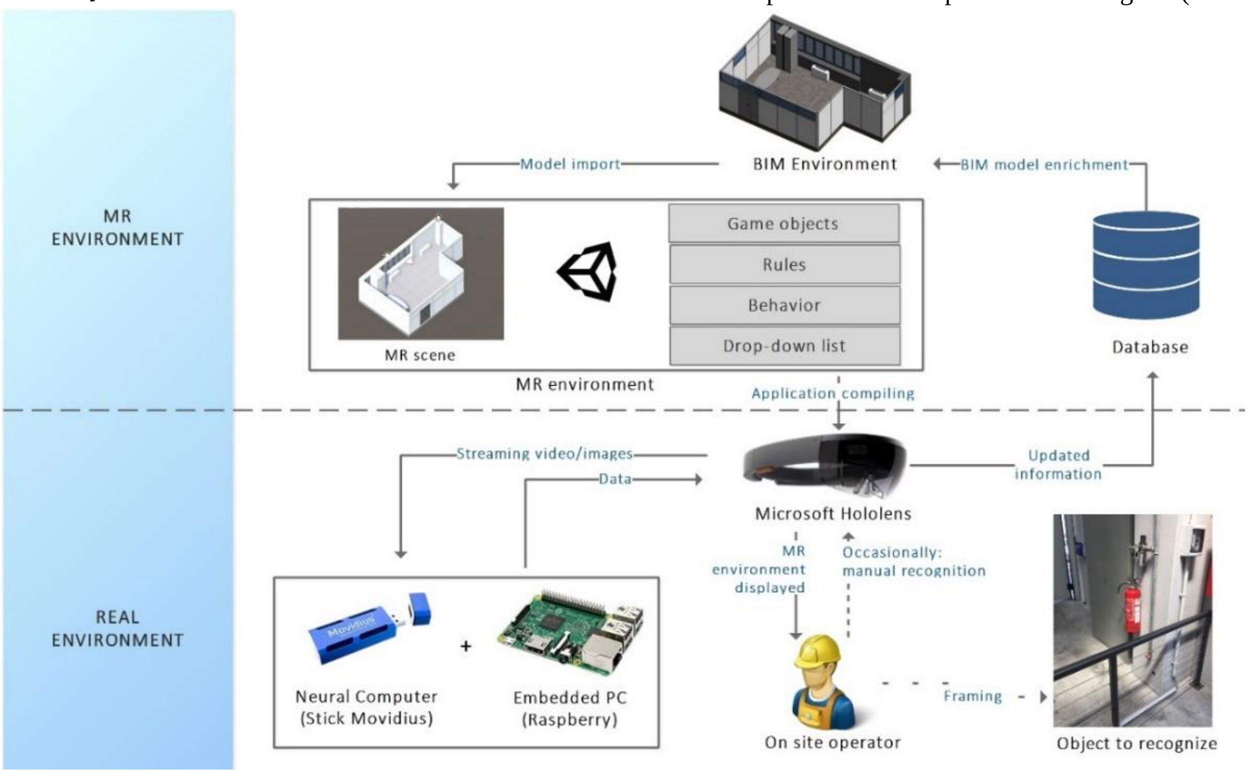

Figure 1. System architecture

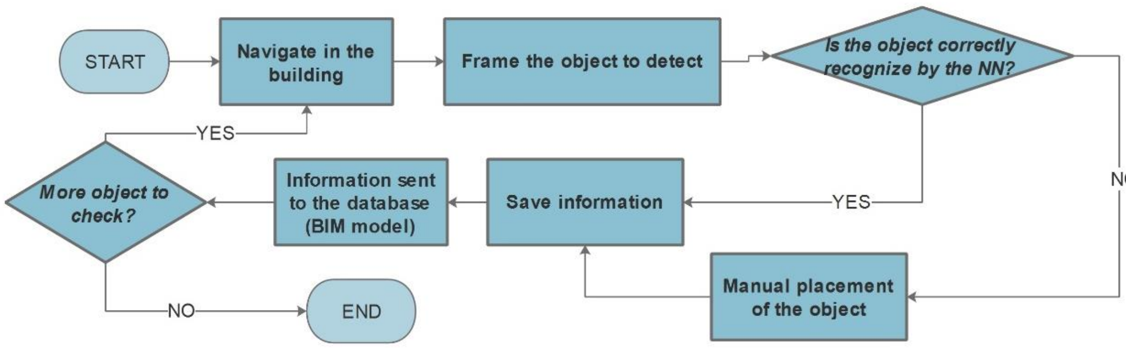

Figure 2. Assets inventory procedure

point the operator is looking at);
2. it sends the streaming video (or images) to the embedded pc;
3. it reads the information provided by the neural network;
4. it acts as interface for the user to manipulate digital objects, their placement in the BIM model.

4 Use cases

The previously explained architecture is specifically thought to be used in two different use cases: the inventory of secondary building assets ; and the support of maintenance operation through asset recognition and retrieval of information concerning selected objects.

These two use cases are discussed in the following paragraphs.

4.1 Assets inventory

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. The system proposed aims to reach a certain level of automation in the process of data collection, especially compared to current inventory procedures that still require long post-processing [8].

For these reasons, the first use case aims to improve the efficiency of survey procedures as well as to lead to the automatic enrichment of BIM models.

The procedure proposed (Figure 2) with this system considers an operator wearing the MR head-mounted display and walking into the building for the first time. Once any asset (secondary component) to detect is framed by the MR device the image is sent in real time to the neural network which is trained to recognize a specific category of objects. The neural network sends back the label associated with the image (as an outcome of the recognition) and the coordinates of the position of the object (assessed from the drawing of a bounding box containing the element). These data allow the application to choose a specific object, among a predefined library, and to place it, as a hologram, into the real environment. In the case of incorrect or missing identification, the operator has still the possibility of manually placing the object, as a hologram, into the real world. All the data about the detected objects and their positions are transferred into a NoSQL Database so as to be available for future operations.

The procedure just explained will enable the operator to check real time the collected data and to modify them if necessary while on site. This will reduce the possibility of errors or of losing. AR will represent a valuable means to display the gathered information so as to interact with them on site. The operator and the application proposed will work in parallel leading to a higher efficiency in contrast with current methods where there is a prevalence of man-machine working in series. This use case has been partially developed as explained in section 5.

4.2 Maintenance operation support

The architecture proposed can also be valuable to support maintenance operations. Buildings contain a series of assets that need maintenance and which are subjected to regular inspection. Correct and immediate localization of objects requires time and it involves a greater possibility of error in complex buildings.

With the objects recognition system the procedure starts with the inspector walking around the facility with the head-mounted display. The application in the device sends continuously in real time the video of what the operator sees to the embedded system. Once the operator frames a component of interest, a pre-trained neural network in the embedded system performs an identification. The application automatically shows the hologram of the object identified, with all its components, and asks for the information to additional show as holograms (e.g. procedures, check-lists, user manuals, usage data). The technician has the possibility to check information needed and to fill in forms at the end of operations. The information entered in the application is transferred to the FM database so as they can be consulted in case of further maintenance operations.

On one hand this procedure will ensure better performance on site, thanks to the possibility to show larger amounts of information on a digital device and directly related to the object. This will save time used to retrieve the useful information. On the other hand, the real time updating of data reduces the time necessary to fill in the database with this information and support the operators in preventing data loss. Furthermore the automatic localization on site is highly beneficial in case of buildings where the operator is working for the first time, or to avoid manual errors (e.g. wrong floor selection). Finally, this system would definitely provide support with new personnel since the MR allows to show detailed step-by-step procedures directly in their field of view.

5 Preliminary development

5.1 From BIM to Unity

Information from the BIM model necessary to create the proper MR scene. First, the geometry of the building is required to allow an accurate positioning of the secondary components. Secondly, component definition, parameters and details are needed to improve the accuracy of the location data. For instance, an object fire extinguisher has to be placed on wall surfaces, thus walls need to be sufficiently detailed. Materials too could be practical as it would make the vision more pleasant and realistic and it would help the operator in identifying surfaces and components.
In order to upload the BIM model into the Unity project the method chosen was to use the *.fbx* format. Importing such file directly into Unity without any other operation results in a loss of graphical information, so it is fundamental to do a second passage through 3D Studio Max so as to re-assign materials to objects. However, with this first method, non-graphical object information is not showed in Unity. The recognition system for inventory requires the geometry more than these non-graphical parameters. However, for the second use case, object parameters are necessary in order to update or consult data during operations. For this reason, further studies will be conducted on a new method to import non-

Figure 3. Secondary element recognition embedded system

graphical along with graphical object information, possibly by deploying a COBIe or IFC data parser.

5.2 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 in the building, and therefore in its digital twin;
2. to send the streaming video 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, 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.

5.3 Streaming video and data transfer Hololens – embedded system

The purpose of this system was to have the whole process performed on site. From the sending of the first piece of information, the video, up to the positioning inside the digital building of the components, in the form of hologram, everything is carried out on site and real time (Figure 3). The communication protocol supporting the information flow for the recognition process from the Hololens to the embedded pc is defined as follows. In order to let the Hololens communicate with the Raspberry, a custom protocol over TCP/UDP is employed. This communication works through a local network and it is able to transfer both streaming video and data deriving from the recognition.

5.4 Neural Network

The recognition of small objects takes place by means of trained neural networks. The objectives of this recognition process are to identify the right type of object (e.g. fire extinguisher size), and locate it accurately within the frame ;

Several types of neural networks exist and the YOLO is the one chosen for this project, for the following reasons [17, 18]:

- the speed which is 45 frames per second; this means streaming video can be processed in real-time, with negligible latency of a few milliseconds.
- the one-step method for classification and localization of objects;
- the simultaneous prediction of multiple bounding boxes;
- the simultaneous prediction of multiple label confidence score;
- it is an open source solution.

In this project, a pre-trained YOLO is used. With this kind of neural networks, it is possible to re-train the last level of the network to detect the object of interest. The network needs a dataset of images to learn how to recognize the object of interest. The dataset features are explained in the following sub-section.

5.4.1 Dataset creation

The dataset to train the network to recognize a specific object should have specific features. With the aim of identifying only the right object, i.e. a fire extinguisher, the dataset will include at least one image for every existing type of fire extinguishers, so as to be able to recognize the fire extinguisher no matter the external appearance. As the dataset must include thousands of pictures, it will be made up of both original pictures and graphically re-edited photos as suggested by studies on dataset creation [19]. The process of editing the original photos is performed with the help of an augmentor software program available online.

The creation of the dataset involves labelling all the images, both with the bounding box around the object to be recognized and with the label assigned to it. YOLO requires 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).

The first goal was to recognize the object but next steps will handle the recognition of different fire extinguisher type. For this purpose, the distinctive components of the fire extinguisher need to be recognized separately. For instance, the pressure gauge helps in the identification of the extinguishing agent, the horn too is helpful in defying the type.

5.4.2 Neural Network training

A first training has been carried out using a tiny YOLOv2 pretrained with the COCO dataset.
YOLO requires some files to start training which are [17, 18]:

- total number of action classes (1 for our case);
- text file with the path to all frames which we want to train;
- text file with names of all action classes (fire extinguisher);
- the path to save trained weight files;
- a configuration file with all layers of YOLO architecture (described in Figure 2).
- pre-trained convolutional weights.

The value of filters in the configuration file of YOLO (.cfg file) for the second last layer is not arbitrary and depends on the total number of classes. The number of filters can be given by: filters=5*(5+number of classes). In order to start training the original .cfg file, the structure must be modified in its following features:

- batch=64, this means we will be using 64 images for every training step;
- subdivision=8, the batch will be divided by 8;
- classes=1, the number of categories we want to detect;
- filters=30, from the previous formula;

Figure 4. Training image type

Figure 5. Drawing the bounding box and labelling the object

- 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. The network automatically creates two groups of images, the former to train the network and the latter to test and validate it.

6 Tests and Results

First tests have been made about the training of the Neural Network. All these tests have been carried out on a computer in this preliminary phase.
The YOLO tiny v2 chosen has been trained with three different datasets in order to check if there was any improvement in the detection of fire extinguisher.
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%) obtained with dataset augmentation;
- 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%) come from the augmentation process.

For every dataset the images were divided into two groups, one for the training and one for the test, with the following splits: first dataset (D1) 260 training images and 40 testing images, second dataset (D1+D2) 429 training images and 71 testing images, third dataset (D1+D2+D3) 590 training images and 110 testing images. 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). In the test performed we worked using Visual Object Tagging

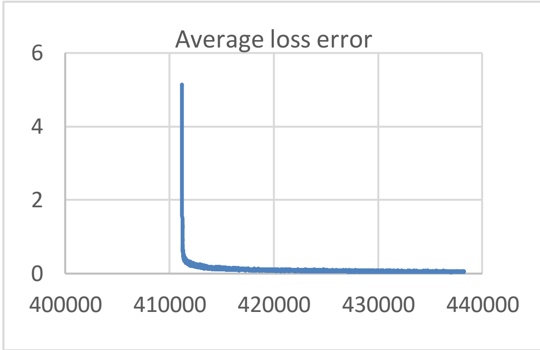

Figure 6. Decreasing of average loss error training 2

Tool for the manual drawing of the bounding box and for the assignment of the label (Figure 5).
The output of this process was a series of .txt file (one for every picture) containing the coordinates of the boxes and the category label (fire extinguisher in this case).
With the first dataset the training was stopped after 22100 iteration. For every training the choice of the weights was done extracting the indicators every 1000 iteration and then choosing the best one in the no longer decreasing area of the average loss error (Figure 6). 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 of 110 photos. The results are displayed in Table 1. 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.

Table 1. Training results

	TR 1	TR 2	TR 3
N° of recognized fire extinguisher	42	44	50
N° of multiplied bounding box for one fire extinguisher	26	16	14

7 Conclusion and next steps

This research aims to provide a support to FM operations during the building lifecycle. The system proposed combines three different technologies: the AR, the BIM data model and Deep Learning, all in one embedded solution for on-site use. The two use cases described could improve their efficiency thanks to the effective Human-Computer Interaction (HCI). First feasibility tests have been conducted related to the training of Neural Network and the recognition through the AR device. Further steps will be developed, starting from the increase of the number of images in the dataset so as to achieve a higher precision in the recognition process. Furthermore, after having reached a higher level of precision the new datasets will be developed for the recognition of precise type of fire extinguisher. Next steps will include also 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.

[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).

[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.

[7] Bonanni, T. M. et al., 2013, Human-Robot collaboration for semantic labeling of the environment, Cognitive Cooperating Robots, pp. 1-25.

[8] 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

[9] 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.

[10] 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.

[11] 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.

[12] Riexinger, Günther, et al., 2018, Mixed Reality for On-Site Self-Instruction and Self-Inspection with Building Information Models. Procedia CIRP, 72: 1124-1129.

[13] 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.

[14] Hugo S., Resende R., Breternitz M., 2018, Mixed reality application to support infrastructure maintenance. In: Young Engineers Forum (YEF-ECE), 2018 International. IEEE, p. 50-54.

[15] 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.

[16] 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

[17] 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.

[18] Shinde, S., Kothari, A., & Gupta, V. (2018). YOLO based Human Action Recognition and Localization. Procedia Computer Science, 133(2018), 831–838.

[19] 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).