Toward automated generation of parametric BIMs based on hybrid video and laser scanning data

Ioannis Brilakis a,*, Manolis Lourakis b, Rafael Sacks c, Silvio Savarese d, Symeon Christodoulou e, Jochen Teizer a, Atefe Makhmalbaf a

a Georgia Institute of Technology, USA
b Foundation for Research and Technology, Greece
c Technion Israel Institute of Technology, Israel
d University of Michigan, USA
e University of Cyprus, Cyprus

Article info
Article history:
Received 21 June 2010
Accepted 24 June 2010
Available online 16 July 2010

Keywords:
Building information modeling
Computer vision
Machine learning
Image processing
Videogrammetry

Abstract
Only very few constructed facilities today have a complete record of as-built information. Despite the growing use of Building Information Modelling and the improvement in as-built records, several more years will be required before guidelines that require as-built data modelling will be implemented for the majority of constructed facilities, and this will still not address the stock of existing buildings. A technical solution for scanning buildings and compiling Building Information Models is needed. However, this is a multidisciplinary problem, requiring expertise in scanning, computer vision and videogrammetry, machine learning, and parametric object modelling. This paper outlines the technical approach proposed by a consortium of researchers that has gathered to tackle the ambitious goal of automating as-built modelling as far as possible. The top level framework of the proposed solution is presented, and each process, input and output is explained, along with the steps needed to validate them. Preliminary experiments on the earlier stages (i.e. processes) of the framework proposed are conducted and results are shown; the work toward implementation of the remainder is ongoing.

1. Introduction
The current state-of-the-art approach to collecting, organizing and integrating as-built data of a constructed facility into a single data structure is to model it using building information modelling (BIM) tools [1]. This approach generates parametric building models by producing logical building objects and the parametric relationships among them. The process starts by collecting spatial data on site through state-of-the-art surveying technologies, such as laser scanning (LIDAR) and photogrammetry. The resulting spatial data must then be manually stitched into a 3D surface with some algorithmic help for fine stitching. The points on the 3D surface are then manually replaced by objects, by having an operator observe the data, identify each object type, search for it in a database of standardized objects, and fit it on the surface with some help from fitting algorithms for optimal fitting. Following that, any as-built attributes can be assigned to each object manually.

Although this process is significantly assisted by recent technological advancements, most of it remains manual. Researchers along with professional modellers such as VECO [2] and Reality Measurements [3] have reported that more than two thirds of the efforts needed to model even simple facilities are spent on manual conversion of the surface data to a BIM. This problem results in significant cost and effort that is needed to convert the sensed surface of constructed facilities to the desired model, which undermines any benefits of automated spatial modelling for the majority of facilities. According to studies by Minhindu and Arayici [4] and Young et al. [5], BIM adoption is growing in some countries such as U.S., Denmark, Finland and Norway. However, as McCarthy [6] had predicted, for small construction projects, the net savings can barely justify adoption and utilization of this technology. As a result, the penetration of innovative spatial modelling technologies to smaller projects and companies in the Architecture, Engineering & Construction (AEC) industry is slow and they will wait unless significant savings can occur.

This paper presents a novel framework that holds promise to automate almost entirely the generation of as-built parametric BIMs of constructed facilities, ranging from residential housing to industrial structures. This framework uses spatial and visual data collected in the field to generate images and the 3D surface represented as a point cloud. The next step is to stitch images together in order to integrate them into a single 3D representation. Then, by analyzing geometric surface and surface texture information...
In an environment with full degree of freedom is a unique feature of laser scanning that distinguishes it over traditional surveying. Furthermore, laser scanners allow for wide range measurements at high resolution and are generally not limited by ambient conditions during operation. Finally, laser scanners have the capability to collect data fast and provide a large amount of information about the surface of a facility in the form of a dense set of 3D points, by simply using a laptop computer, a laser scanner, and a tripod. The quality of the surfaces generated by this method is relatively high. The generated surfaces are precise and detailed enough to be used in fine modelling applications such as elevation spatial modelling of soil and obstacle tracking, according to Hashshash et al. and Kim et al., respectively. However, Kim et al. and Caldas et al. asserted that these methods can take many hours or even days to model a structure.

The only information the sensed 3D surfaces generated from the methods presented above contain is spatial measurements. The difference between the 3D surface generated and the as-built object-oriented model is that the surface does not have any semantics corresponding to the elements it contains or any information pertaining to the relationship of an element and its adjacent ones. Also, it cannot provide any other as-built information such as material information and structural health. So, it should be converted into an information-rich, object-oriented model which contains multiple elements with a wide range of attributes such as material, schedule and cost. The process of transforming a 3D surface (spatial data) to an object-oriented model involves several procedures. Using modelling tools (such as Autodesk’s Revit and Grasshopper’s ArchiCAD), a human modeller can recognize building elements and crop them from the 3D surface. These pieces should be matched, fitted and stitched together to be included in the generated 3D model. The relationships between elements and their adjacent objects can be defined automatically. Non-spatial related attributes, parameters and property definitions are then added to the model manually. Performing these procedures is currently almost entirely human dependent, which makes the conversion of sensed infrastructure surface to the object-oriented model time-consuming and costly.

3. Object recognition and fitting

The framework proposed in this paper is partially based upon the object recognition and fitting procedures presented in this section. Object recognition in as-built spatial modelling is the task of identifying the elements that are related to the construction of facilities. According to Brilakis and Soibelman, object recognition in construction is influenced by characteristics of construction material images, which have low variability and high similarity (e.g., wood, concrete, steel, and earth) as opposed to generic databases (e.g., face, cars and planes). As a result, generic methods developed in other fields cannot satisfy recognition of objects in construction related applications. There are different methods developed for recognition of construction objects by Kwon et al., Bosche and Haas, and Brilakis et al. Material-based and shape-based recognition are two recent techniques that are outlined below.
In material-based object recognition, features such as color, texture and structure are used to identify and classify object elements in images. Material-based recognition consists of three steps which are object representation, matching and classification [18,19]. The first step, which is object representation, is performed by decomposition of each image into its basic features such as color, texture and structure, by applying filtering techniques. The next step, which is material matching, can be done by cropping the image into regions using clustering methods [19]. This is followed by computation of the material signatures of each cluster. Meaningful image clusters should then be identified and isolated by comparing each cluster signature with the feature signatures of the materials in a model database that stores material image samples. The final step is image classification, which basically classifies images into similar classes of objects, shapes or materials depending on the type of classification we are interested in [51]. For example, in material-based recognition, this is performed based on the material information extracted from them (i.e. construction material in this case) and stored in a knowledge-based system [19].

Shape-based recognition is a complementary technique which uses shapes for automatic identification of linear and nonlinear construction objects within the image content. Similar to other object recognition methods, shape recognition consists of three steps, which are shape representation, shape matching and shape classification [52]. There are different features that can be used to represent the shape of an object such as the silhouette, the points, or the skeleton. One method that was implemented and tested for construction-related applications is mainly based on a skeletonization algorithm [15]. This algorithms works by selecting objects (image areas) with similar features and a certain degree of uniformity, recognizing the materials that consist of each cluster, computing the maximum cluster dimension (MCD) and the maximum dimension along the perpendicular axis of MCD (PMCD) (Fig. 2), and finally using these dimensions to evaluate the linearity and orientation of the “object’s spine” for each cluster [15]. If MCD is a lot larger than PMCD, the object is linear and its direction on the image plane can be computed based on the tangent of the MCD edge points. It can also be determined if the object is a column or beam if the direction computed is within 45° from the vertical or horizontal image axis accordingly. There are techniques developed by Kwon et al. [16] for fitting and matching which take advantage of the fact that most construction objects can be fitted and matched using a few types of geometric shapes. For example, cuboids, cylinders and planar surfaces can be used to fit and match objects such as columns, pipes and walls accordingly.

Furthermore, as shown by Zhu et al. [20], in detecting structural components such as concrete columns typically one should look at (a) shapes dominated by long straight vertical lines and (b) uniform texture and color patterns. As such, first detected in a video frame are the long vertical line pairs. The process is facilitated by a Canny edge detector [21] which is used to produce an edge map which is then used in a Hough transform [22] for retrieving the long vertical-line information in the edge map. Each retrieved vertical line is then compared to its neighboring vertical lines (if two vertical lines have similar size then they are regarded as one pair of vertical lines) in an iterative comparison process. When all pairs of long vertical lines are found, a bounding rectangle for each pair of them is constructed, the texture contained in the bounding rectangle is retrieved and then compared with the signatures of traditional concrete textures. If indeed the texture is matched to concrete then the region within the pair of long vertical lines is taken to be the boundary of one concrete structural element with long vertical lines [20]. Concrete columns and walls are examples of such concrete-based items, while beams are not since they do not have long vertical lines. The ratio of the width over the length of the remaining rectangles is calculated and if the value is small the object is classified as a concrete column [20].

4. Proposed framework

The proposed framework aims to automate the generation of parametric BIMs of facilities in terms of their spatial and visual features. It starts by sensing the real-world structure and ends with a BIM. The basic mechanics of this framework are shown in Fig. 3. There are six states (inputs/outputs) in the proposed model generation process, which are represented by elliptical shapes in Fig. 3 while processes that are needed to carry one state to the next are shown in boxes. The processes are discussed below.

4.1. Visual and spatial sensing

Visual and spatial sensing is the first process in this framework. The input to this process is the real world structure that is captured by laser scanners calibrated with cameras. Its outputs are the images and the 3D surface represented as a point cloud. This recon-
struction process of a scene in 3D space by using image sequences to measure and determine the coordinates of objects in 3D is referred to as videogrammetry [53]. In order to cover large structures, several range scans need to be obtained from different viewpoints so that field-of-view and self-occlusion limitations are overcome. These scans should then be integrated into a single representation corresponding to a coherent 3D model of the real environment, which gives rise to the registration problem. This concerns the task of computing a 3D rigid transformation that aligns together pairs of range datasets that have been scanned from different locations. Depending on the amount of relative displacement and orientation between range scans, the cases of fine and crude alignment are differentiated. Fine alignment assumes roughly registered datasets and can be reliably carried out automatically using the Iterative Closest Point algorithm [23] and its robust variants to iteratively calculate point correspondences. On the other hand, crude registration is a much more challenging and, therefore, still open problem. This stems from the requirement of consistent performance even when there is very little overlap between datasets. To achieve the rough alignment necessary to bootstrap ICP, contemporary commercial systems rely on either manual matches specified via interactive user input or fiducial scene marks, calling for a time-consuming and laborious process.

In the proposed work, the emphasis is on fully automating the crude alignment case, avoiding the use of external coordinates (e.g. GPS) or environment modifications. Several directions towards this end will be explored. One possibility is to employ the camera that is bundled with the range scanner and in order to obtain combined 3D point measurements and geometry-registered intensity images. Then, SIFT keypoints and related viewpoint invariant descriptors [40] that are known to be highly distinctive representations of local texture [24] will be extracted and matched between the intensity images. At least five such matched keypoint pairs suffice to estimate the relative camera motion between the two intensity images using structure from motion vision techniques [25]. Embedded in a robust regression framework [26], larger numbers of matches can provide increased robustness. Knowledge of the relative position between the range scanner and the camera allows the estimated camera motion to be converted into a rigid displacement achieving the crude alignment of range datasets. Multiple datasets can be aligned simultaneously with a global refinement step using bundle adjustment techniques [27]. Another possibility is to base crude alignment on the putative point correspondences obtained with the aid of localized surface shape representations provided by spin images [28] or Thirt features [29]. A third possible line of work is to employ extended Gaussian images (EGIs) and to describe shapes by mapping a scan onto a sphere using the estimated surface normals. Recently, Makadia et al. [30] reduced the EGI to a “constellation” of peak locations on the Gaussian sphere and matched these constellations between scans. This matching yields several relative pose hypotheses between scans; each of which is then validated geometrically and the best one is selected for achieving crude alignment.

4.2. Spatial correlation

Spatial Correlation is the next process in the proposed framework. The inputs to this process are the aligned point clouds and the corresponding calibrated intensity images obtained by registering several scans; the output is a textured 3D surface model describing the overall topology of scanned structures. The surface model is generated as a manifold approximating the point cloud with simple geometric elements such as triangles, a process also known as meshing or grid generation. Meshing is necessary for reducing geometric complexity and adapting the range data to the requirements of computer graphics software. Optimal triangulations that are best according to some criterion that measures the size, shape, or number of simplices can be computed [33]. However, a simple approach where a 3D mesh is generated by back-projecting a 2D triangular image mesh according to the depth values obtained via scanning, is often sufficient. A desirable set of properties for integration algorithms includes robustness to gross outliers, smoothness of the resulting mesh (but not at the expense of destroying fine details and discontinuities) and amenability to numerical implementations that are efficient, incremental and yield a globally optimal result. In [34] the desired manifold is extracted as the zero isosurface of a cumulative function constructed from signed distance functions. More recent approaches apply a certain level of regularization in order to favour smooth surfaces [35].

Knowledge of the exact geometric relationship among range and intensity images permits the latter to be employed as texture maps by simply defining the texture coordinates using the 2D coordinates of vertices. Median filtering of texture values can discard imaging defects like sensor noise, specular reflections and highlights, thus improving the quality of textures. Another possibility for enhancement stems from the fact that the observation of a point on an object’s surface with multiple images of limited pixel resolution gives rise to subpixel shifts in each available image. This can be exploited to construct a super-resolution texture by fusing all images on a finer resampling grid, thus increasing detail and overcoming aliasing. It is also important to note that not all images are equally appropriate for texture extraction since the object distance and viewing angle have a considerable impact on texture appearance. Therefore, when extracting the texture of a certain surface region, the images depicting it under favourable conditions should be given precedence. Texture quality also depends upon the characteristics of the employed image acquisition hardware. For instance, the imaged intensity of a certain object may vary considerably across a set of images due to the auto exposure feature included in most cameras for compensating for their limited dynamic range. Such exposure differences are likely to give rise to visible seams when combining textures originating from different images, thus calling for radiometric calibration of images. Visible seams can also arise due to white balance variations or geometry misregistration errors. Techniques for creating clean texture maps by intelligently deciding which pixels to use and how to weight or blend them are discussed in [36]. As a final post-processing step, texture manipulation operations such as inpainting can be employed to remove unwanted surface marks.

4.3. Object features recognition

Object Features Recognition is the next process which handles recognition of the objects’ visible attributes using image processing tools. The input to this step is the rendered 3D surface that contains the correlated spatial and visual data of the structure to be modelled. The output is the augmented 3D surface containing recognized visual and spatial features of potential objects as well as the background. This process is concerned with the identification and selection of most appropriate visual and spatial features, selection of a suitable algorithm that most adequately represent the distinctive pattern of each feature, and finally, incorporating a novel method that can make feature representation invariant with respect to variations of viewpoint and illumination conditions. To satisfy these concerns, robustness of image analysis tools in recognition of information-rich features and selection of appropriate algorithm for feature detection and description that also handles invariance are critical. Recognition performance of this step is influenced by data accuracy and density of the input, which should be determined by validating the outcome of the previous steps.
As Savarese and Fei-Fei have shown in [37], compact models of an object category can be obtained by linking together diagnostic parts (also called canonical parts) of the object from different viewing points. As previous research has shown, a part-based representation [38,39] is more stable for capturing appearance variability across instances of objects. Our canonical parts are discriminative regions of the objects that are comprised of many local vector quantized features [40,41]. Such parts retain the appearance of a region across instances of objects. Our canonical parts are discriminative parts (also called canonical parts) of the object from different views of the object's appearance across multiple views. Instead of recovering a full 3D geometry [42,43], canonical parts are connected through their mutual homographic transformations and positions. This information is captured by pose matrix $H_{ij} = [A_{ij} t_{ij}]$, where $A_{ij}$ and $t_{ij}$ are the homography and the translation vector representing the relative change of pose and position, respectively, between canonical parts $i$ and $j$. The resulting model is a compact summarization of both the appearance and geometry information of the target. Unlike earlier attempts for 3D object categorization, our framework requires minimal supervision and provides the capacity to estimate an object's pose along with its class label. Because the object is represented as a linkage structure of parts, recognition is robust in the presence of partially occluded objects; even if one or multiple parts are occluded, other parts can still gather further evidence for the target. We can employ this framework for accurately detecting and recognizing objects from arbitrary poses and under arbitrary illumination conditions in a highly cluttered environment.

We use the “Car” category as an example to better define this concept. Fig. 4a shows a car within the viewing sphere. As the observer moves on the viewing sphere the same part produces different appearances. The location on the viewing sphere where the part is viewed the most frontally gives rise to a canonical part. The appearance of such canonical part is highlighted in yellow. In Fig. 4b, colored markers indicate locations of other canonical parts. Canonical parts are connected together in a linkage structure as shown in Fig. 4c. The linkage indicates the relative position and change of pose of one canonical part given another (if they are both visible at the same time). The changes of location and pose are represented by a translation vector $t_{ij}$ and a homographic transformation $A_{ij}$, respectively.

### 4.4. Object classification and size fitting

Object classification and size fitting are applied next. The inputs to this task are the recognized visual and spatial features of potential objects and the output is the cropped 3D surface with recommendations on which class these objects belong to. This process is concerned with the characterization and matching of an object with another one from a set of templates in order to determine the type or class of that object. Classification estimates the possibility that an object fits into some specific defined class by determining the distance of that object description to an object classification model, i.e. a template [56]. Based on the semantic labels given to the objects, the available templates in the database are used to match the objects. This can be referred to as progressing from an abstract definition of the object to a more realistic one. For example, a generic wall with brick patterns and spanning between two columns will be classified as a non-load bearing brick wall. A template could be defined by a set of features or using a model [54]. Proper feature selection basically improves the performance of the classifier; it also provides faster and more cost effective classification prediction [55]. Matching is then achieved by estimating the distance of the object description to the template. The type of classifier and the extracted object features used for classification are important factors that improve the accuracy of object classification [55].

To accomplish this process of the framework proposed, it is important to select proper algorithms that most effectively represent the classification of objects, and also, to incorporate a novel method that can make feature matching and classification invariant to viewpoint and illumination conditions variations present in construction sites. Fig. 5 shows an example of object classification for concrete column. The main concern of object classification is the spatial relationships of the object features that together with the recognition of the features themselves make the recognition model.

A combination of artificial neural network (ANN) and fuzzy logic (FL) techniques will be implemented, with the former being good in pattern recognition and rule inference, and the latter being suitable for estimating class posterior probabilities for pattern recognition tasks [46]. An entropy-maximization approach along with entropy’s principal properties of subadditivity and maximality will also be used in finding the global optimum [47] and thus the best object-class match for an object based on the object’s set of features [46], as these features were previously prescribed and identified. Starting with a generic building model database of known objects derived from the IFC schema and a trained dataset (using ANN), inferences can be made using FL for objects newly ‘read’ from the images/video presented to the system. Furthermore, given a set of predefined object features (such as texture, color, shape and material type) and a set of corresponding object-matching probabilities based on the ANN/FL analysis, a neurofuzzy system coupled with entropy-maximization algorithms can then sift through the image data and suggest the best possible class membership for the object in question.

Topological relationships, which describe the relationship between objects by proximity, adjacency, membership, connectivity, enclosure and orientation, are important factors that can significantly improve the recognition performance. Fig. 6 shows a representation of topological relationships in terms of nodes, arcs and
polygons, which form a hierarchy of topological elements to construct topological structures [57]. Relationships such as the relative position of the knob on a door, or the windows of a wall can for example be very helpful in detecting doors, windows and walls. The objects will be represented in an object schema that defines all of the possible inheritance, aggregation and other relationships for each given building or infrastructure system type. Within the developed object schemas, the system will apply a set of topological rules to check whether the initial object identification is reasonable in terms of the physical juxtaposition of the objects. Topological rules will be defined and implemented based on the spatial relationships that are most important for the assembly of building elements. Topological rules can be categorized as line rules, polygon rules and point rules. For example, one of the rules of polygons must be that “they cannot overlap.” This rule requires that the interior of polygons in one feature class does not overlap with another one. This is useful for the classification of two mutually exclusive systems of area classification, such as walls and windows (areas defined as windows cannot thus also be defined as walls, and vice versa). An example of topological rules and their application can be found in Borrmann and Rank’s work [58] on the topological analysis of 3D building models using a spatial query language, in which basic topological operators are examined and used in reflecting the topological relationships between 3D spatial objects. Borrmann and Rank’s work presents definitions of the semantics of such topological operators and describes a possible implementation of the topological operators by means of an octree-based algorithm which enables users the handling of topological relationships in a fuzzy manner [58].

For more accurate classification, all possible configurations of objects from the probabilistic data provided will be tested. Where relationships inferred from the relative locations of objects are not feasible, some configurations and some possible object type identifications will be removed, thus narrowing the uncertainty in the identification process. Once the object is finalized, the rules use the list of attributes required to fully represent the object. If the object is for example classified as a “Column”, then the building elements hierarchy is searched to find the exact node that contains the representation for column. The property sets, entity-relationships and metadata required for completing the data set, are retrieved this way at a specified level of detail. Size fitting is used to extract the exact model with the same size and dimension as the column found.

### 4.5. Human-assisted model assembly

Human-assisted model assembly is the next process involved. While fully automated modeling is a goal to be pursued, it is highly unlikely to be achieved economically in practice. The purpose is to automate the process as far as possible in order to make the job of the modeller easier. The complexity of the problem is such that interim results for the preceding stages will only find application in the short term if human assistance is facilitated.

Thus we consider the possibility that if any ambiguity remains after the first step, human assistance will be sought. Using the available geo-referenced video frames overlaid in a suitable BIM system interface with the objects identified automatically, an operator can quickly identify errors in the model proposed by the system. Four types of correction are likely to be needed:

1. **Completion of missing objects.** Where the automated routines fail to identify an object, the user will model the object in the same way as they would place a new object in any BIM application, i.e. by selection from the native object schema, setting local parameter values and inserting the object in the correct location and with the correct orientation. Selecting location and orientation will be facilitated by snapping to the scanned geometry.

2. **Correction or replacement.** Where an object has been wrongly identified, the operator will substitute the correct object type for the incorrect one.

3. **Aggregation or unification.** There may be situations where objects appear as sets of divided objects due to occlusion at specific points along their length (such as a single continuous partition wall identified as a series of segments because it was scanned in a series of separate rooms). An operator will be able to union such objects. Similarly, where implicit relationships exist, such as in a row of identical windows, an operator will be able to aggregate them.

4. **Separation.** In the inverse case, where a single object has been identified but it is in reality an aggregation of joined objects, an operator could quickly separate objects as needed (for example, where a structural column identified as a single vertical object might in reality be composed of multiple joined sections).

This review process will be carried out using a customized interface built within a standard BIM application which incorporates the range of possible building object types in its internal object schema. The operator’s visual inspection will be enhanced by color-coding the background geometry according to confidence levels generated in the previous step to highlight where objects may be missing or possibly inappropriate.

### 4.6. Parameterization and application of constraints

In this final stage, the parameters needed for the internal representation of each object will be estimated, each according to their geometry class according to Eastman et al. [11] (chapter 3). The result of the previous stages is a set of 3D objects that have been classified as building components. The geometry will include objects in two forms, dependent on their source:
5. Results

Preliminary experiments have been conducted for the first three processes of the framework proposed (i.e. visual and spatial sensing, spatial correlation, and object features recognition) and results are obtained. In an initial investigation, a number of laboratory setups have already been developed for the evaluation of laser scanning capabilities and for comparing the accuracy of photo-generated vs. laser-generated point clouds. Three experiments have been conducted to obtain data from different environmental settings, as both photographs and laser scanning data are affected by their surroundings.

5.1. Sensing and spatial correlation

The experimental object used for 3D surface reconstruction is a single object cuboidal masonry block (scanned both indoor and outdoor). By scanning the same object both indoor and outdoor, the accuracy differentials can be calculated. The object was scanned so that all visible faces were captured including sides and top (if possible). For each masonry block, a 3"x3" target was affixed to the sides and top to be used in post-processing for scene reconstruction and determining three-dimensional accuracy. This method is more robust than previous research efforts from El-Omari and Moselhi [31] that attempted to combine laser scans and photos from one face of an object. In addition, since the scanned objects were at actual construction sites, the data collection is more realistic and representative of scenarios faced by surveyors.

Between 50 and 200 images were taken for each experimental object. Images were taken at a high speed exposure level to aid in fast photo image capture and for each object a circular path was traversed, with images being shot at approximately every meter. Depending on the size and location of the object, a different number of pictures was needed. A commercially available hi-resolution laser scanner (Leica ScanStation 2) was used to obtain laser range point data of the same scene. For each object, a three-point survey traversal was used with scan data taken from each location. As in traditional surveying, each scan station includes both a fore-sight (following scan point) and backsight (previous scan point). Each location produces a standalone individual point cloud of a scene that, when combined with other point clouds from multiple field-of-views, can create a true 3D point cloud of a scene and its objects. For each experiment, both hi-resolution and low resolution laser scans were performed at each station. Once the three experimental objects were scanned, the photo-log was taken and point clouds were registered using commercially available software (Leica Geosystem Inc. Cyclone 6.0 and Cloudworx). The D4AR system developed by Golparvar-Fard et al. [32] was utilized for visualizing the results.

Registration of the scanned points took less than 10 s for each scene. Manual extraction of the 3D laser point cloud of the masonry block from the point clouds overall returned minimal noise and thus the only modification to the model involved deleting points that were not on the masonry block (i.e. the surrounding environments). Next, photo images taken by the laser scanner were overlaid on the point clouds in order to create better realism. A semi-automated step selects the remaining points of the masonry block and forms a CAD object. This is done using a proprietary algorithm created by Leica Geosystems Inc, which can currently automatically apply simple geometric shapes, such as columns, blocks, and pipes if enough points are present. Three different views of the returned laser scanned data for the masonry block used in the experiment are shown in Fig. 7. Fig. 7(a) shows a digital photo of the masonry block. Targets have been placed on the block to aid in the calibration and accuracy measurements. Fig. 7(b) shows a high density laser scan of the block and the resulting CAD object produced, while Fig. 7(c) shows the block with a photo overlay from the integrated camera of the laser scanner and Fig. 7(d) represents the idealized lines.
5.2. Features recognition

The experimental object used for object features recognition is a structural concrete column on a construction site. Artificial intelligence techniques have already been utilized by Zhu and Brilakis [48] for detecting concrete regions in images, by first employing image segmentation to divide an image into several regions and then extracting the visual features of each region and processing them with an artificial neural network. The artificial neural network, which is originally trained with a database of images, helps identify whether a region in an image is composed of concrete or not. This way, all concrete regions within an image can be identified (Fig. 8), without the necessity of predefining sample color values for concrete region identification as other methods require.

6. Conclusion and ongoing research

Having access to an as-built model of an existing facility can enhance project planning, improve data management, support decision makings and increase the productivity, profitability and accuracy of a project in construction industry. However, current building modelling methods for existing buildings rely heavily on human effort to generate the logical building objects and the parametric relationships between them. Data can be collected automatically using laser scanners, but interpretation of point clouds, stitching and object fitting are all performed manually. This process is time-consuming and costly, which counteracts the benefits of as-built modelling for civil infrastructure projects.

The framework proposed in this paper can achieve the integration and employment of ‘intelligent’ components and methods that can support a wide range of structures, thus achieving automation of the modelling processes involved in the generation of BIM. Successful implementation of this research is expected to significantly reduce the time and resources needed to model a structure, making as-built modelling affordable for even smaller construction projects. This can spearhead the infiltration of as-built modelling in the Architecture, Engineering, Construction and Facilities Management (AEC/FM) industry, for both design and construction. Construction in developed countries involves a lot of renovation and expansion of existing facilities: access to cheaply and automatically compiled parametric as-built models would enable designers to use BIM tools without rebuilding the existing structures. In construction, many practical applications have been developed in research, but they have not been applied in practice because of the cost of as-built modelling. These applications include quality control, and productivity and performance measurement according to Akinci and Boukamp [49] and Navon and Sacks [50], respectively, and various forms of construction automation that use guidance systems.

Ongoing research work builds on knowledge already acquired by the authors and previously reported on [1,15,18–20,32–33,45,47–48] and entails the implementation and validation of the independent hypotheses behind each of the processes proposed in the framework. Each process will be implemented and the output(s) of that activity should be validated before it can be used as an input to the next process. The implementation and validation of the proposed framework poses several unique challenges that require research at the intersection of BIM, visual and spatial sensing and sensor systems [59], computer vision and image processing technologies. This work faces several challenges that are aggravated by the need to deploy and utilize the expected prototype in harsh, outdoor environments. The main challenges involved in this investigation is the integration of video capture and laser scanning technologies to capture not only the precise geometry but also generating semantically meaningful components of structures such as columns, pedestals, arches, vaults and

---

**Fig. 7.** (a) Actual image of masonry block, (b) returned point cloud over fitted CAD object, (c) point cloud with masonry block images overlaid, and (d) idealized lines.

**Fig. 8.** An example of concrete region detection [48].
beams. Another challenge is to create a set of parametric building
objects to represent the components, including their behaviour,
their connections and the relationships between them such as
aggregation and structural support. The ultimate goal of this work
is to provide models that can be imported into the native object
forms of commercial BIM tools so that they can be manipulated
to represent the components, including their behaviour,
beams. Another challenge is to create a set of parametric building
objects to represent the components, including their behaviour,
beams. Another challenge is to create a set of parametric building
objects to represent the components, including their behaviour,


