Automation in Construction 96 (2018) 211–223

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Prioritizing object types for modelling existing industrial facilities

Eva Agapaki[a,*], Graham Miatt[b], Ioannis Brilakis[a]

[a] Department of Engineering, University of Cambridge, United Kingdom
[b] AVEVA Group plc., United Kingdom

ARTICLE INFO

Keywords:
Building information modelling
Existing industrial facilities
Modelling time
Automated modelling
Facility management
Scan-to-BIM

ABSTRACT

The cost of modelling existing industrial facilities currently counteracts the benefits these models provide. 90% of the modelling cost is spent on converting point cloud data to 3D models due to the sheer number of Industrial Objects (IOs) of each plant. Hence, cost reduction is only possible by automating modelling. However, automatically classifying millions of IOs is a very hard classification problem due to the very large number of classes and the strong similarities between them. This paper tackles this challenge by (1) discovering the most frequent IOs and (2) measuring the man-hours required for modelling them in a state of the art software, EdgeWise. This allows to measure (a) the Total Labor Hours (TLH) spent per object type and (b) the performance of EdgeWise. We discovered that pipes, electrical conduit and circular hollow sections require 80% of the TLH needed to build the plant model. We showed that EdgeWise achieves cylinder detection with 75% recall and 62% precision. This paper is the first to discover the most laborious to model IOs and the first to evaluate state-of-the-art industrial modelling software. These findings help in better understanding the problem and serve as the foundation for researchers who are interested in solving it.

1. Introduction

"As-Is" Building Information Models (AI-BIMs) are the 3D digital representation of the existing condition of facilities and encompass geometric definitions at different levels of aggregation and parametric rules [1]. The clear majority of large refineries were built before the advent of CAD in 1977: as-is models, therefore, do not exist to assist their maintenance operations [2,3]. AI-BIMs of industrial plants have substantial impact in various applications. Some of these include maintenance, strategic planning of their operations, revamping purposes, retrofitting of old sites and preparation for dismantling [4–7].

Inexistence of AI-BIMs will result in time lags for these operations. This is crucial for industrial managers, since without detailed planning, productivity will be substantially affected, and the agreed budget and timeline expectations will not be met. Moreover, there are thresholds on the acceptable shut down duration that will not impede production, and those limits cannot be violated without incurring extra costs. For instance, Sanders [45] reported that 40% of the total 3D modelling cost of retrofitting a Chevron plant was spent on data-processing labor and the shut-down time was limited to 72 h to avoid additional costs. Every modelling hour saved can prevent critical failures or unexpected accidents, thus continuous production flow of these assets is achieved. This work aims to assist the tedious current practice in this regard.

Modelers use the following four main steps to manually process AI-BIMs: (a) data collection, (b) point cloud registration, (c) geometric modelling and (d) addition of accompanying information. Initially, data is collected using laser scanners and photogrammetry, which are represented by their Cartesian or polar coordinates, the point cloud, and in some cases by their color data (RGB). The scans need to be registered in a consistent coordinate system by calculating inter-scan rigid body transformations and the registered point cloud represents the complete measured data. Then this data needs to be geometrically modelled.

Geometric modelling entails (a) primitive shape detection, (b) semantic classification of detected shapes and (c) fitting. Firstly, primitive shapes are detected (e.g., cylinders, tori, planes) and classified (e.g., pipes, elbows, I-beams). Afterwards, the primitives are fitted to known solid shapes to obtain their geometric parameters. Their relationships to other objects need to be obtained in order to produce a complete AI-BIM in the Industry Foundation Schema (IFC) format. The IFC schema is a software-agnostic platform that allows geometric, material and other construction related information to coexist in a single model.

Geometric modelling is the "bottleneck" during the Scan-to-BIM modelling process of any industrial facility given how costly and time consuming it is. Recent studies have reported that geometric processing takes 90% of the modelling time [8,9]. Hullo et al. [9] reported that 10 operators were needed to process 1084 scans of a nuclear reactor and

* Corresponding author.
E-mail addresses: ea437@cam.ac.uk (E. Agapaki), Graham.Miatt@aveva.com (G. Miatt), ib304@cam.ac.uk (I. Brilakis).

https://doi.org/10.1016/j.autcon.2018.09.011
Received 22 January 2018; Received in revised form 20 June 2018; Accepted 19 September 2018

model its objects in around 6 months using Dassault Systems SolidWorks and Trimble Realworks. In contrast, laser scanning of the plant was completed in only 35 days. This significant time required to model the large number of industrial objects impedes adoption of as-is 3D modelling for these plants.

The research presented in this paper is exploratory in nature, not causal. It does not seek to solve the problem of automating the modelling of industrial facilities. It rather seeks to improve our understanding of the problem and the extent to which it has been resolved so far and provide a foundation for future researchers interested in solving it. This is why the main objective of this paper is to identify the most important industrial object types given how frequent and laborious they are for modelling, as well as to measure the performance of existing tools in modelling these particular object types. The authors identified the most frequent objects based on a frequency-based, statistical analysis of 3D modelled industrial objects in a variety of industrial plants. The most frequent objects were then modelled in the state-of-the-art, semi-automated modelling software, EdgeWise, and their modelling time was measured. Finally, the most important industrial object types were ranked based on their frequency of appearance and average modelling time. This analysis will substantially assist automated modelling efforts to efficiently reduce modelling time and facilitate facility management.

2. Background

Industrial plants can be divided into fifteen main categories [10]: (a) onshore and (b) offshore oil platforms, (c) chemical, (d) mining, (e) pharmaceutical plants, (f) power plants, (g) water and wastewater treatment facilities, (h) natural gas processing and biochemical plants, (i) refineries, (j) food processing factories, (k) defense facilities, (l) metal production facilities, (m) nuclear plants, (n) research facilities and (o) warehouses and silos. The object types of industrial facilities belong to the main object categories: (a) structural elements, (b) piping system, (c) electrical, (d) safety and (e) general equipment, (f) architectural elements, (g) instrumentation, (h) Heating, Ventilation and Air Conditioning (HVAC) and (i) civil elements. Representative examples of structural elements include barricades, catwalks, mod pilings, steel platforms, stairs, pipe racks, supports and structural steel elements. Respectively, examples of safety equipment include deluge systems, cameras, fire extinguishers, fire aid stations and fire detectors. General equipment includes lifting mechanisms, pumps, compressors, tanks, turbines, vessels, degassers, air coolers, drainers, water heat recovery units and exchangers. Civil elements include curbing, foundations and bollards. Examples of architectural elements are windows, slabs and walls. Instrumentation includes sensors (temperature, pressure, etc.) and controllers. Indicative examples of electrical equipment are cable trays, conduit, electrical panels, power outlets and lights.

2.1. Value of modelling industrial object types

Petitjean [11] prove that 85% of objects in industrial scenes can be approximated by planes, spheres, cones and cylinders. These primitive shapes, however, have not been assigned to specific industrial object types. The value of modelling those is measured in terms of safety, maintenance and retrofitting [12]. AI-BIMs for industrial plants have significant value for facility managers since these models assist them to be proactive in decision making that involves maintenance, operations and health and safety. Recent studies of the Chartered Institute of Building [13] have shown that the need for refurbishing and retrofitting 93% of existing industrial facilities will be a major focus in the U.K. construction industry by 2050. As a result, modelling these assets using digitization technologies is an imperative need.

Extensive research has been conducted to identify critical industrial objects under the above-mentioned values of modelling [14–20]. Susceptibility to failure is measured based on failure rate metrics. The nominal mean failure rate (λ_0) is the frequency that an industrial object type or object component fails and is usually expressed in failures per year [17]. The sample data for electrical component failures can be combined from different data sources and calculation of a mean failure rate is reasonable. Moss and Strutt [17] list several factors that affect the mean failure rate of mechanical components in industrial facilities. These factors depict the design, the size of equipment, environmental conditions and level of operation compared to the mechanical capacity of an object [17]. For example, outdoor facilities that are affected by more challenging weather conditions tend to be more prone to rust. The same paper specifies factors calculated to modify the standardized life of a component given those factors. Particularly for chemical plants and offshore platforms, these factors increase the nominal mean failure rates of mechanical components due to environmental conditions and heavy equipment operation compared to average industrial conditions. Steel sections are also critical for fatigue and fire, dependent on the load imposed and welding [14,15].

The criticality of industrial object types is then defined as the likelihood of failure multiplied by the consequence of failure for an industrial object or a process line of a plant [19]. There are three methods in literature used to evaluate the hazards and assess the consequences of accidents for a plant. These are HAzard and OPerability (HAZOP), Failure Mode and Effect Analysis (FMEA) [16] and Fault Tree Analysis (FTA) [18]. What is missing, though, is a justified study on which critical objects should be modelled for maintenance, safety or retrofit purposes.

Examples of critical object types that should be considered are given below. Hazardous subsystems should be modelled in finer detail for safety purposes. Highly hazardous object types are separators, compressors, driers and flash drums, whereas moderately hazardous ones are pipelines and pumps [20]. The identification of hazardous equipment elements will remarkably improve safety management.

Valves are a final control element in nearly all chemical process control loops and regulate the flow through piping systems. Failure to quickly locate and identify control and safety valves during inspection can result in significant damages or even massive, unprecedented disasters such as Texas City Refinery [21] or Piper Alpha [22]. Safety system deficiencies that occurred due to poor inspection and inadequate maintenance are reported as some of the main factors of the devastating incidents mentioned above.

Another important control measure in industrial facilities is maintenance of pipelines and pipe supports. Insulated pipes and pipelines carrying flammable, hazardous or toxic materials are highly important for inspection. One of the most important concerns of inspectors for maintenance of pipelines is corrosion. Pipes of Nominal Bore (NB) > 2 in. (50 mm) are considered critical for corrosion [23].

Structural steelwork and equipment are also vital for the structural stability of the plant and oil and gas production especially in cases of fire. Given the short lifecycles of refineries, which range from 15 to 30 years, structural design is challenging since the layout should be flexible and expandable [24]. Seismic and energy refurbishments for pipes are typical retrofitting operations in industrial plants [25]. AI-BIMs can significantly assist these operations, should accurate as-is models of these objects be created.

Table 1 summarizes the critical elements for each category (maintenance, safety and retrofit) based on their failure rates λ_0 (high, medium and lower impact) based on Umar [20] and Keeley et al. [26]. These values are calculated for major accidents that involve dangerous substances and cause serious damage/harm to people and/or the environment. The piping system is generally subdivided in two meaningful subgroups with respect to their Outer Diameter (OD). Small bore pipes are the pipes whose OD is less than or equal to 2 in. (50.8 mm) and the rest (pipes with OD > 2 in.) are considered large bore pipes. Table 1 shows that small bore pipelines are considered to have higher impact than large bore. Some categories listed in Table 1 are critical but not frequent. For this reason, they do not appear in Tables 3–5.

The critical industrial object types have been investigated in the literature. However, those that need automated modelling due to

Table 1
Critical object type list for facility management in terms of value for modelling.

Value for modelling	High impact ($\lambda_0 \geq 10^{-4}$) yr^{-1}	Medium impact ($10^{-5} \leq \lambda_0 \leq 10^{-4}$) yr^{-1}	Lower impact ($\lambda_0 \leq 10^{-5}$) yr^{-1}
Maintenance	Valves	Small bore straight pipes	Large bore straight pipes
Safety	Separators, reciprocating compressors, driers & flash drums, valves, large vessels, tanks, electrical conduit, circuit breakers	3 mm diameter straight pipes, pumps, reciprocating compressors	4 mm diameter straight pipes, 25 mm diameter straight pipes, 33 mm diameter straight pipes, pressure & spherical vessels
Retrofit	–	3 mm diameter straight pipes	4 mm diameter straight pipes, 25 mm diameter straight pipes, 33 mm diameter straight pipes

increased modelling time and frequency of appearance have not been identified. If an object type is critical but not frequent, there is no need to automatically model it. On the other hand, even if an object is valuable for modelling but is not frequent, this paper does not consider it for automated modelling. The primary condition that the industrial objects should meet in order to be considered for automated modelling is being in the priority list based on their frequency of appearance.

2.2. Frequency based studies

There is no substantive study that prioritizes industrial objects based on their frequency of appearance as explained in Section 2.1, however there are related fields where object importance is considered for object classification [27,28]. SceneParse150 [28] is an image dataset, part of ADE20K, used for image classification that contains the eight most frequent object classes ('person', 'building', 'car', 'chair', 'table', 'sofa', 'bed', 'lamp') and 150 objects in these classes found in a variety of everyday scenes. The uniqueness of this dataset compared to other benchmark datasets, such as ImageNet [29] and Pascal [30], is that the distribution of objects that appear in the selected images is diverse, which mimics object occurrences in daily scenes. This dataset, however, is limited to everyday scenes and not extended to industrial facilities. Therefore, the statistics of most frequent object types in industrial scenes are not determined. As such, the identification of the most frequent object types in industrial plants facilitates the application of multi-classifiers and makes a difficult multi-classification problem solvable. Then, the researchers can focus their detection efforts on the most frequent object categories that take most of the manual labor time, so that users manually model those that take less modelling time. Application of the results of the frequency based studies will guide researchers on automatically detecting and classifying these objects in industrial scenes. A training library of the object classes that are critical for industrial facility operations, frequent in industrial plants and laborious to model will assist the implementation of multi-classifiers for automated modelling of these classes.

2.3. Automated industrial plant modelling

2.3.1. State-of-the-art software

The next challenge for plant modelling is that almost all available modelling tools depend on human intervention for most of the modelling tasks. Leading 3D CAD software (Autodesk, Bentley, AVEVA and FARO) have developed programs containing a variety of functions that enable pipe modelling from 3D point clouds. Automated detection has been achieved by a limited number of software packages. For example, AutoCAD Plant 3D accompanied with FARO's PointSense Plant add-in enables semi-automated pipe modelling from Point Clouds. PointSense Plant provides several functions and a large standard library with a variety of piping and structural components available for the detection of pipelines from 3D point clouds. Moreover, fitting template objects to scanned 3D objects is performed automatically and constraints can be applied to fix potential errors of fitting. PointSense Plant 17.5 has integrated a pre-calculation tool that detects cylinders in the point cloud of a specific area and has the ability to colorise the Point Cloud by deviation from reference geometry [31]. However, the users still manually model the as-is pipelines by finding the insertion points for fitting CAD objects to the segmented 3D point clouds and fitting errors of the extracted cylinders are not provided. The "*Walk the Run*" feature is rather a suggestion for pipe insertion points than an automated pipe modelling tool.

EdgeWise is another semi-automated platform that is extensively used. The main difference of Pointsense and EdgeWise is that a modeler using the former should manually extract the desired boundaries of an object and afterwards the software will automatically extract the correct dimensions and location. However, this procedure is automatically performed by EdgeWise, that is why it was chosen as the most suitable tool for evaluation of the most frequent industrial object types that will be presented in Section 3.4. Structural sections are manually modelled in all available software packages. Fitting of user-selected primitives (e.g., circular hollow sections, cuboids, tori etc.) is performed automatically by both EdgeWise and PointSense Plant. To date, no one has provided viable and accurate assessments of state-of-the-art tools.

2.3.2. State-of-research

State-of-the-art research work on pipe detection has partially solved the problem and not to a greater extent compared to commercially available software like EdgeWise [32–34]. For instance, Ahmed et al. [32] only detect pipes in orthogonal directions. A recent study completed by Patil et al. [33] is dependent on threshold values for radius and normal estimation. The pipe radius range is 0.0254 m–0.762 m and the normal deviation is 5°. Therefore, Patil et al.'s study cannot be generalized for pipe detection. Their updated Hough Transform based on Rabbani et al. [5] detects pipes in two sample datasets with 60% recall and 89% precision. Sharif et al. [34] proposed a model-based cylindrical and structural object detection by matching features of the acquired point cloud data with those of library generated point cloud models. However, the experiments are limited to a small-scale pipe spool and a structural frame.

Prior knowledge of industrial scenes has assisted researchers to detect industrial objects. Son et al. [6] used prior knowledge (Piping and Instrumentation Diagram, P&ID) to detect Mechanical, Electrical and Plumbing equipment (MEP). However, as-is P&IDs are often not available as prior knowledge in industrial plants, thus they do not reflect the modifications a plant undergoes through its life. For this reason, prior knowledge cannot rely on P&IDs. Perez-Gallardo et al. [35] used topological information to extract semantic labels for four object classes: pipes, planes, elbows and valves. They detect cylinders with 86% precision and 92% recall. However, their semantic labels consider that all cylindrical objects are pipes, without investigating other potential object classes with the same shape.

3. Research methodology

3.1. Gaps in knowledge and research questions

Considering the state of practice and body of research reviewed above, existing studies for as-is modelling of industrial plants have primarily focused on automated detection of cylindrical objects and no scientific and viable evaluation of existing state-of-the-art software

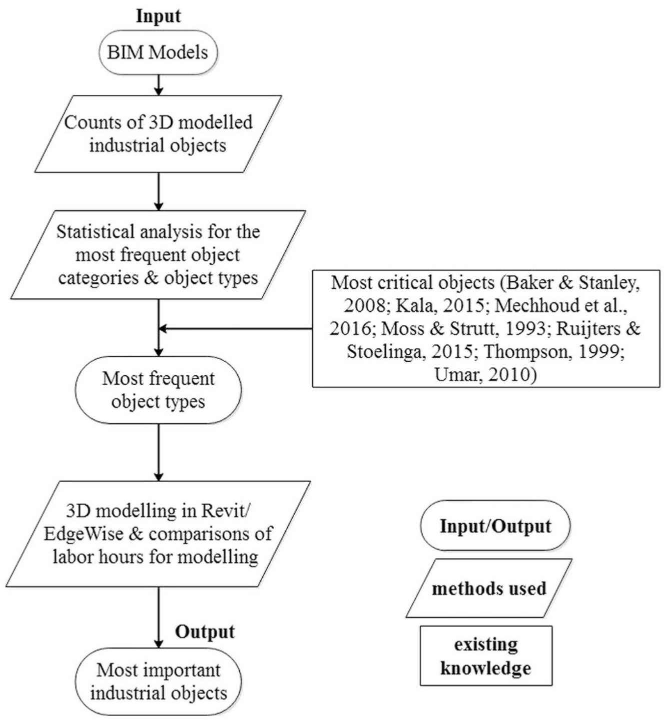

Fig. 1. Research methodology.

tools is provided. Critical industrial object types have been identified in the literature based on their value for modelling, but no scientific study investigates modelling those. It is therefore still unclear (1) which industrial object types are important for automated modelling, (2) how long it takes to model those in state of the art software and (3) the level of automation achieved with state of the art software.

The aim of this work is to solve the gaps in knowledge by answering the following research questions:

a) What are the most important industrial object types in terms of value for modelling, frequency of appearance and modelling time?
b) What is the time required for modelling the most frequent object types in state-of-the-art software?
c) How can state-of-the-art as-is modelling tools be assessed in terms of automated detection of objects achieved?

The research conducted in this paper is exploratory in nature and follows the methodology framework depicted in Fig. 1. We analyzed the counts of 3D modelled industrial objects obtained from as-designed BIMs by hierarchically ordering those based on their average frequency of appearance in sample case studies. The most frequent object types were then modelled in EdgeWise to measure the modelling time of each type. We determined a list of the most important object types as those being most frequent and most laborious to model. The time required for manual modelling of cylindrical objects was then compared with that measured in EdgeWise.

3.2. Data collection and assumptions

Five case studies of 3D modelled industrial facilities were examined to find a statistically representative sample of object types in industrial facilities. Three case studies were offshore platforms, one was a petrochemical plant and the fifth was a food processing refinery (sugar refinery). The subcategories of offshore platforms that were examined in this study are (a) a Gravity-Based Structure (GBS), (b) a Tension-Leg Platform (TLP) and (c) a fixed platform. These facilities are anonymized since rights are reserved by AVEVA Group plc. and British Petroleum (BP). The total number of objects in the sugar refinery and petrochemical plant is 22,143 and 240,687 objects respectively. The GBS, TLP and fixed platform have 577,237, 434,780 and 34,089 objects respectively.

An assumption was made for electrical, safety equipment, HVAC and civil categories of the offshore platforms and sugar refinery due to unavailability of data. This assumption for electrical and HVAC categories is reasonable, since the pipe network and fittings can be simulated with conduit and valves/flanges. The percentages of these categories for the petrochemical plant were used to calculate the respective percentages in the other case studies presented. For instance, in the case of electrical equipment, around 27% of the total objects in the petrochemical plant was assumed to be present in the tension-leg platform, which has 289,943 objects in structural, piping system, equipment, architectural and instrumentation categories. These categories represent 67% of the total objects in the facility, which are 434,780 in this case, assuming that safety objects constitute around 6%, HVAC 0.6% and Civil 0.05%. The same concept is applied to identify the object counts in the other missing categories of the case studies.

This study is not seeking to find the statistics of industrial object categories and object types by decimal accuracy. It rather proposes the most frequent object categories and object types that need to be prioritized for automated modelling. For this purpose, the datasets were chosen to represent different types of facilities with total number of objects that differs by orders of magnitude. As such, it is legitimate to assume that the ranking of object categories would not be substantially different if additional facilities were assessed.

The safety equipment (deluge systems, cameras, fire extinguishers, fire aid stations and fire detectors), civil elements (curbing, foundations, bollards) and HVAC were not modelled in the majority of the as-designed BIM models that we investigated (four out of five as-designed models as described above). Safety equipment, civil elements and HVAC are approximately 6% of the total number of objects in the petrochemical plant and this assumption was adopted for the other datasets as well.

Electrical equipment is substantial when modelled (27% of the total number of modelled industrial objects in the petrochemical plant). We observed the correlation between the average frequency of appearance and the total number of objects in the categories where the statistics were available for all the investigated case studies. A sensitivity analysis on the total number of industrial objects that a plant can have, is conducted to observe the range of percentages for each object category. Fig. 2 shows the average curves on the data available from the counts of as-designed BIM models of industrial plants.

The results shown in Fig. 2 indicate that the prioritization of the object categories does not change with the increase in the total number of objects. As such, it is reasonable to assume that when electrical equipment is modelled, its hierarchical order compared to the other categories will not change between facilities. For cases where modelers ignore modelling electrical equipment, this does not indicate absence of electrical equipment in the existing industrial plant.

The range of total number of objects is defined based on the existing datasets ($1.5 * 10^4$–$6 * 10^5$ objects). The average frequency of appearance based on our data is a linear function for object categories shown in Fig. 2, object types of the pipe system and structural elements in Fig. 3(a) and (b). Equipment, architectural elements and instrumentation are < 5% in all case studies.

These observations substantiate the assumptions for the average frequencies of the categories with unknown data, since the average frequency of appearance is invariant to changes in the priority list of objects. This means that the prioritization of object categories does not change with the increase in total number of objects except the piping system and structural elements. These two categories overlap at 240,200 objects and their hierarchical order is reversed for greater number of objects above this threshold. We observe that piping elements and structural elements vary from 20 to 40%, given our data. The

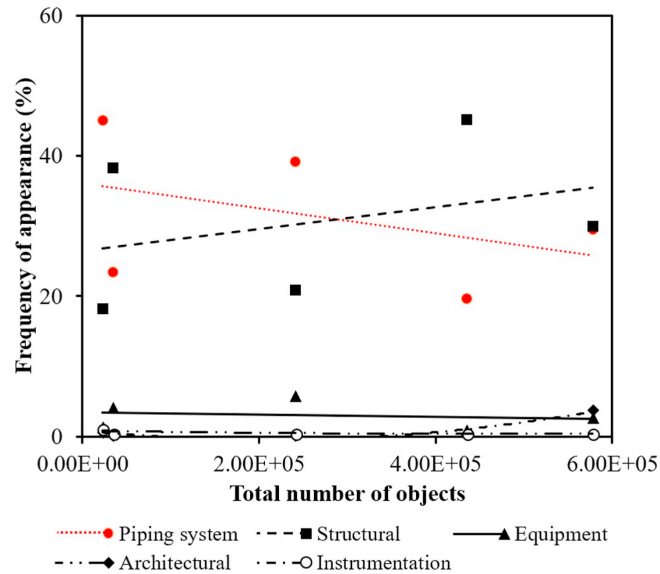

Fig. 2. Estimated frequency of appearance (%) of industrial object categories with respect to the total number of objects in the five case studies of industrial plants.

variance in frequency of appearance is justifiable given that the variability in the total number of objects is more than one order of magnitude. The correlation between the total number of objects and the prioritization of object categories is not significant despite the variance in frequency. This means that the hierarchical order of the most frequent object types does not change with the change in total number of industrial objects with the exception of piping and structural elements as explained.

No correlation between the size of the plant and the frequency of appearance is observed. There is no clear definition of the size of a plant compared to its total number of objects. The results on object categories show a decreasing trend with increasing total number of objects, except structural elements that have an increasing trend. The same trend is observed for all object types other than solid bars and I-beams, whose average frequency increases with the total number of objects.

3.3. Most frequent industrial object categories

The object categories that need to be modelled are determined by implementing a statistical analysis on the frequency of appearance of all object categories encountered in typical industrial plants. The frequency of appearance is calculated by dividing the total counts of each object category ($n_i = S_i$, P_i, E_i etc. dependent on the category) with the total number of objects of the same object category in all case studies (N_i). The total counts of each object category were calculated by running a Programmable Macro Language (PML) script in 3D models designed in Everything 3D [36] software. The pseudocode of this script is shown in Algorithm 1 for finding the counts of 3D modelled pipes, structural elements and equipment. The counts of the rest of the object types are calculated in the same manner. The pipe length corresponds to the length of the pre-fabricated pipe spools that are manufactured offsite and are supplied in the following cut lengths: single random and double random [37]. The former is usually 16–20 ft on average, whereas the latter comes in lengths of 35–40 ft on average [37,38].

Algorithm 1. Pseudocode used to obtain number of objects from 3D modelled industrial plants.

Input: 3D modelled elements (pipes, structural and equipment) {P, S, E}

Output: Number of pipes {P_i} for each bore, number of structural elements {S_i} for each type, total number of equipment {E_i}

For each pipe P

- Traverse the pipelines and store each pipe length L_i, where L_i represents the length of each pipe spool
- Collate the total number of pipe spools on each element type of pipe bore P_i

End For

For each structural element S

- Determine the structural element profile S_i
- Calculate the structural element length and store for each profile type SL_i

End For

For each equipment element E

- List the element description E_i and store for each element

End For

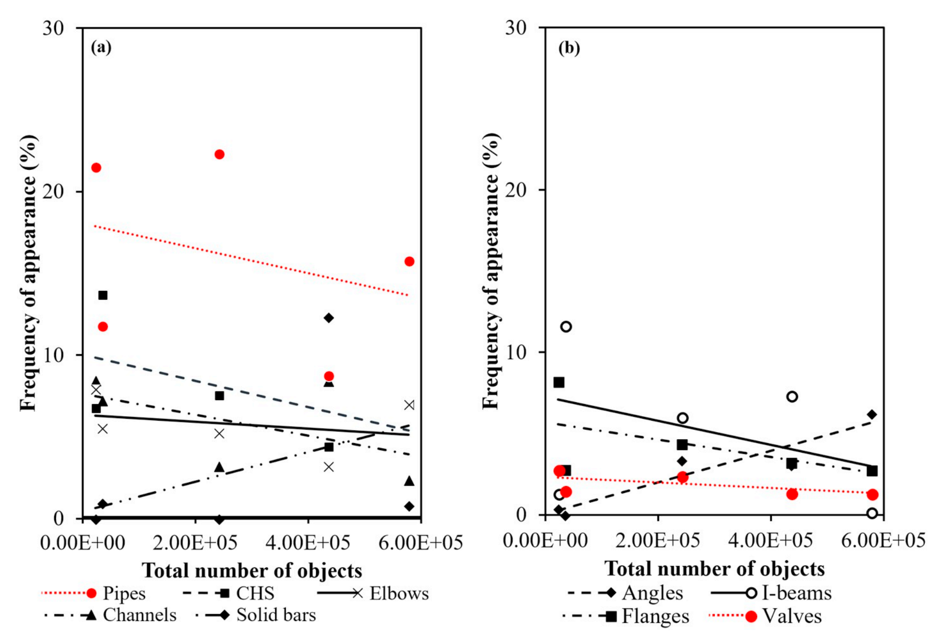

Fig. 3. (a) & (b) Estimated frequencies of appearance (%) of industrial object types with respect to the total number of objects in the five case studies of industrial plants.

Subsequently, the most frequent object types of these categories are presented. The standard error of each object category between the five case studies using the inter-project standard error (S.E.) is given below [39]:

$$S.\,E._{\cdot 1-5} = z\sqrt{\sum_{k=1}^{5}\frac{p_k\,(1-p_k)}{n_k}} \quad (1)$$

where $k\,\epsilon$ [1;5] for each individual case study, p_k is the probability of appearance of each object category in case study k, n_k is the number of objects in each category and z is the Z-score corresponding to the confidence level of a Gaussian distribution.

The object category rankings are calculated in descending order for all case studies in Table 2. Structural elements are most frequent in all case studies with an average frequency of around 33%. The piping system and electrical equipment follow in percentages being 28 and 27% respectively. These statistics are important since most software packages and research methods are designed to automate only the modelling of pipelines, electrical conduit and Circular Hollow Sections (CHSs), which are all cylindrical objects. Each object category follows a binomial distribution where N is the total number of objects in all independent object categories existing in our case studies (1,308,936 in all five projects) and p_i the probability of appearance of the specific object category i. A binomial distribution can be approximated to a Gaussian distribution if the following conditions are met [40]:

$$N_i > 30 \quad (2)$$

$$N_i p_i > 5 \quad (3)$$

$$N_i(1-p_i) > 5 \quad (4)$$

The results show that all conditions are met for every object category, thus the approximation to a Gaussian distribution is valid. The sample size of the binomial distribution of each object category, the standard deviation and standard error of the sample mean are also presented in Table 2. The sample size of the object category i is defined as [39]:

$$n_i = N_i p_i \quad (5)$$

where p_i is the frequency of appearance of the object category $i\,\epsilon$ [1;9], since there are nine independent object categories for each case study.

The standard deviation and standard error of each object category i are calculated using the equations for a binomial distribution ([39,46]):

$$\sigma_i = \sqrt{Np_i\,(1-p_i)} \quad (6)$$

$$S.\,E._{\cdot i} = \frac{z\cdot\sigma_i}{\sqrt{n_i}} \quad (7)$$

The standard error estimates the standard deviation of the sample mean based on the population mean. This definition implies the sample follows a Gaussian distribution [39] as proved above. The standard deviation for the structural category is 540 objects, meaning that there is a higher variance from the mean of all projects (437,530 ± 540 objects) compared to the other object categories, but is low compared to the magnitude of the mean. The standard error with 95% confidence level is low for all object categories, meaning that the sample mean is close to the population mean (1,308,936 objects) for all object categories. For instance, if more samples of each object category are considered, there is 95% confidence level that the average frequency of appearance will be the same as calculated herein. The inter-project standard error, as shown in Table 2 is almost negligible for all object categories implying that the variability of object counts between different case studies for the same object category is very low.

Table 2
Priority list of object categories for all case studies.

Object category	Frequency of appearance (average) (%)	Sample size (n)	Standard deviation (average)	Standard error (95% confidence level)
Structural	33.40	437,530	540	0.92
Piping	28.20	368,428	515	0.88
Electrical	26.90	352,170	507	0.87
Safety	5.70	74,860	266	0.46
Equipment	2.80	36,310	188	0.32
Architectural	2.00	24,557	160	0.27
HVAC	0.60	8431	81	0.14
Instrumentation	0.50	6066	88	0.15
Civil	0.04	584	21	0.04

3.4. Most frequent industrial object types

The most frequent object categories being around 90% of all objects modelled in these facilities are: structural elements, the piping system and electrical equipment. The object types present in these categories were further investigated. Tables 3, 4 and 5 show the priority list of object types belonging to these categories with the same statistical properties evaluated for the object categories.

The results show that Circular Hollow Sections (CHSs) are the most frequent structural elements present in these studies with an average percentage of around 19% (Table 3). They are one of the object types in this category with highest standard deviation (261 objects) and inter-project standard error ($\sim 1.8*10^{-2}$), meaning that their distribution among the five case studies is quite widespread from the sample mean (84,688 objects) compared to the other object types. However, the standard deviation is two scales of magnitude lower than the sample size (n) meaning that the average frequency of appearance is invariant to the sample size, and the samples are large enough to give accurate results. Channel sections, solid bars and I-beams follow with approximately 14, 13.5 and 13% respectively.

The priority list of piping elements is also provided in Table 4. Straight pipes are more than half of the total objects in this category (52.1%) with a slightly higher standard deviation compared to structural elements (303 objects). Elbows and flanges follow with 19% and 12% respectively and lower standard deviations.

Electrical equipment is mostly comprised of conduit (90.2%) in the petrochemical plant as shown in Table 5. An assumption was made that the proportion of electrical equipment in each project will be the same for all case studies as discussed above, thus the inter-project standard error is zero.

The standard error with 95% confidence level for the electrical equipment has the highest range compared to all the other categories, indicating the different scale of total numbers of objects in the five case studies investigated. The standard deviation of the sample considered is

Table 3
Priority list of objects for all case studies in structural element categories.

Structural object type	Frequency of appearance (average) (%)	Sample size (n)	Standard deviation (average)	Standard error (95% confidence level)	Inter-project standard error
CHS[a]	19.4	84,688	261	0.8	$1.77*10^{-2}$
Channel	14.3	62,634	232	0.7	$1.72*10^{-2}$
Solid bar	13.5	58,934	226	0.7	$0.34*10^{-2}$
I-beam	13.1	57,314	223	0.7	$0.58*10^{-2}$
Angle	11.9	51,886	214	0.6	$1.2*10^{-2}$
Others	10.8	47,273	205	0.6	$0.98*10^{-2}$
RHS[b]	9.2	40,439	192	0.6	$0.2*10^{-2}$
PFT[c]	7.5	32,699	174	0.5	$0.28*10^{-2}$
T-brace	0.4	1663	41	0.1	$0.16*10^{-2}$

[a] Circular Hollow Section (CHS).
[b] Rectangular Hollow Section (RHS).
[c] Parallel Flanged Tee (PFT).

Table 4
Priority list of objects for all case studies in piping element categories.

Piping object type	Frequency of appearance (average) (%)	Sample size (n)	Standard deviation	Standard error (95% confidence level)	Inter-project standard error
Straight pipe	52.1	192,081	303	1.0	$1.56 * 10^{-2}$
Elbow	19.3	70,945	239	0.8	$1.25 * 10^{-2}$
Flange	11.8	43,308	195	0.6	$1.09 * 10^{-2}$
Tee & Olet	6.1	22,460	145	0.5	$0.81 * 10^{-2}$
Valve	5.6	20,591	139	0.4	$0.76 * 10^{-2}$
Other	2.2	8137	89	0.3	$0.18 * 10^{-2}$
Reducer	1.8	6570	80	0.3	$0.36 * 10^{-2}$
Cap	1.2	4336	65	0.2	$0.17 * 10^{-2}$

Table 5
Priority list of electrical object types for all case studies.

Electrical object type	Frequency of appearance (average) (%)	Sample size (n)	Standard deviation	Standard error (95% confidence level)
Conduit	90.2	317,572	177	$29.8 * 10^{-2}$
Cable tray	6.1	21,585	142	$23.9 * 10^{-2}$
Electrical panel	2.1	7397	85	$14.3 * 10^{-2}$
Lights	1.4	5009	70	$11.8 * 10^{-2}$
Miscellaneous	0.07	250	16	$2.7 * 10^{-2}$
Alarm	0.05	190	14	$2.3 * 10^{-2}$
Speaker	0.03	120	11	$1.8 * 10^{-2}$
Others	0.01	39	6	$1.07 * 10^{-2}$
Power outlet	0.003	11	3	$0.53 * 10^{-2}$

also low compared to the sample size in terms of order of magnitude, so the average frequency of appearance is a reliable estimation.

This statistical analysis gives us the most frequent object types in object categories that are among the most critical object types for modelling industrial plants as shown in Table 1. According to the results, the most frequent categories are structural elements, the piping system and electrical equipment. These categories represent around 90% of all objects in industrial facilities and the sample sizes for each object category are large enough to give representative results. The most frequent object types of these categories are in descending order: electrical conduit, straight pipes, circular hollow sections, elbows, channels, solid bars, I-beams, angles, flanges and valves. The results are presented in Table 6. It is noteworthy that the rest of the object types present in our datasets were < 1% of the total number of objects, thus neglected from our analysis.

Fig. 4 shows a distribution of ranked object types with their corresponding average frequencies for the five case studies investigated. The distribution follows the Zipf's law [41] and is typically found in everyday scenes as explained in Section 2.2. This means that the average number of industrial objects and their ranking are inversely proportional. Therefore, the most frequent object category (electrical conduit) will occur approximately twice compared to the second most frequent category (straight pipes), three times as often as the third most frequent category (CHSs) and so forth. The 10 rank-ordered object types can be used for automated modelling. This industrial object repository can then be the dataset used for training machine learning algorithms. Automated modelling of those categories will significantly assist the tedious modelers' work by efficiently reducing modelling time, whereas modelers can intervene to a small subset of infrequent object types.

3.5. EdgeWise evaluation for pipeline modelling

Four sample point cloud datasets were used to evaluate the level of

Table 6
Ranking of most frequent industrial object types of all object categories.

Rank	Most frequent object types	Frequency of appearance (average) (%)
1	Electrical conduit	24.3
2	Straight pipes	14.7
3	CHSs[a]	6.5
4	Elbows	5.4
5	Channels	5.0
6	Solid bars	4.5
7	I-beams	4.4
8	Angles	4.0
9	Flanges	3.3
10	Valves	2.0

[a] CHSs = Circular Hollow Sections.

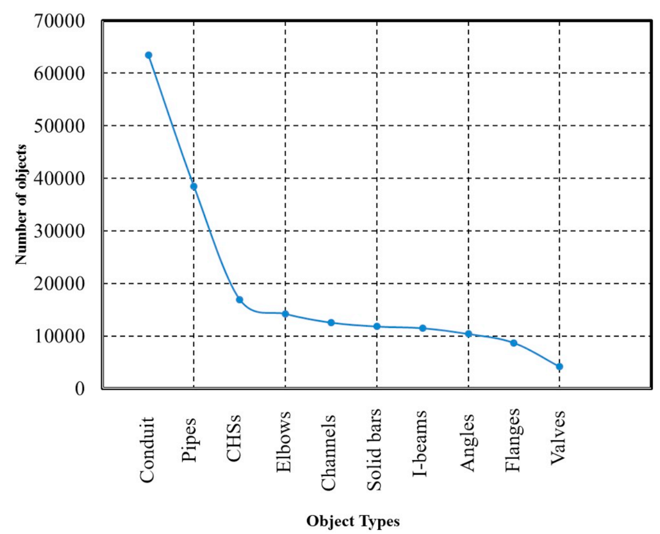

Fig. 4. Object types sorted by frequency of appearance in average number of objects.

automation of EdgeWise and obtain modelling times for the most frequent object types that were identified in Section 3.4. Fig. 5 shows the sample datasets that were used for this evaluation. Two case studies are rooms of an industrial facility, one was a water treatment facility in Cambridge (U.K.) and the fourth was a room of a petrochemical plant. The industrial and petrochemical plants are anonymized since rights are reserved by AVEVA Group Plc. The water treatment facility in Cambridge was laser scanned by the authors.

Pipeline modelling is significantly assisted by the automated extraction of cylinders that EdgeWise provides. The scans were processed on a desktop computer with CPU Intel® Core™ i7-4790 K at 4.00GHz, 32 GB RAM and Windows 10 64-bit operating system. The average processing time for this operation using the above-mentioned operating system for the sample datasets is $3.3 * 10^{-3}$ min/(cylinder * points in the point cloud), as shown in Table 9. The average number of points of all datasets used is 258 million and the number of points of each dataset is presented in Table 7. The average diameter of cylinders and pipes is presented for evaluation purposes in Table 8.

We set the parameters used for cylinder extraction to a minimum of 80 points, in order to detect a pipe and provide a distance tolerance to $0.7 * 10^{-3}$ m. The minimum threshold of the software is 50 points to identify pipelines, however if we give a very low value, the automated extraction tool will identify noisy and erroneous features as pipes. The distance tolerance is a parameter that determines how far away from the cylinder a 3D point can be, so that it is not excluded from the extraction algorithms. The default value of $0.7 * 10^{-3}$ m is used here, which was obtained from a scanner with a high level of accuracy and low noise [42].

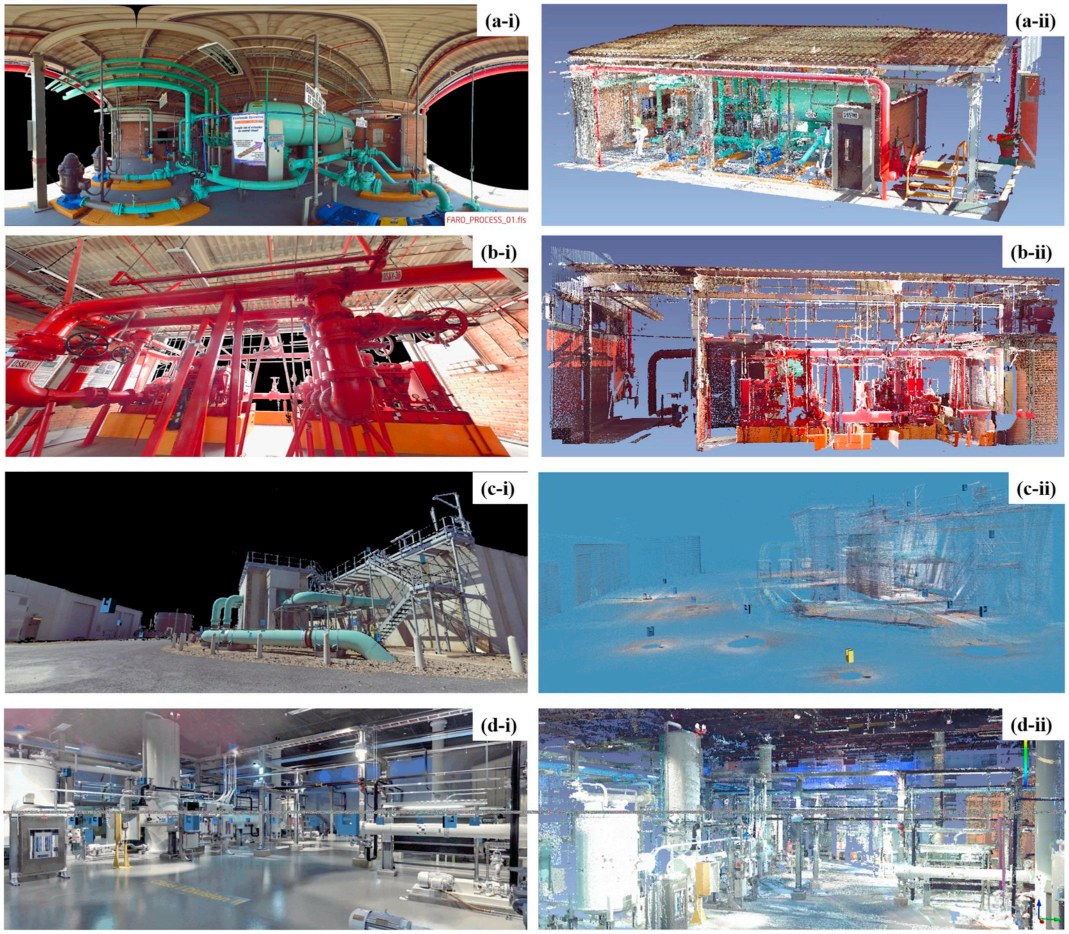

Fig. 5. (i) Evaluated point cloud datasets and (ii) their corresponding sample pictures. (a), (b) Two rooms of a typical industrial facility, (c) water treatment plant and (d) a room of a petrochemical plant.

Table 7
Total number of points in the point cloud datasets, cylinders and pipes in each case study.

	Typical facility Room 1	Typical facility Room 2	Water facility	Petrochemical plant
	Total number			
Points (millions)	129	105	122	675
Automatically detected cylinders	551	86	44	358
Manually detected pipes	166	79	48	265

Table 8
Average diameter of cylinders and pipes for each dataset.

	Typical facility Room 1	Typical facility Room 2	Water facility	Petrochemical plant
	Average diameter (m)			
Cylinder	0.067	0.076	0.315	0.095
Pipe	0.114	0.106	0.617	0.081

After the automated extraction step, the cylinders were inspected and approved depending on the modeler's discretion. For cases where it was difficult to identify the object, pictures taken from the laser scanner were used to assist the inspection process. A user friendly "*Smart Sheet*" was produced, which contains information such as the length, diameter, Root Mean Square Error (RMSE) and coverage (%) of each pipe spool. The results show that although cylinders are automatically extracted, no contextual information is provided. Henceforth, electrical conduit, handrails, cylindrical pipe supports, vessels and other object types were modelled as straight pipes.

The next step in the evaluation process was to edit the pipes and to manually add missing ones (classification). Using the "*Easy Connect*" tool pipe spools were connected, and tees and elbows were added in the piping network. Then, labels were manually assigned for each cylinder that was automatically extracted by the software and metrics were used to evaluate the software's performance.

An additional step of cleaning the pipes and merging the connecting spools together was performed to complete the pipeline system. This step was completed automatically by the software. Then, standard catalogues were used to get standardized pipe dimensions.

We chose the American Society of Mechanical Engineers' (ASME) specifications and pressure rating of 150 psi. After this step, fittings, such as flanges and valves, were applied on the standardized pipes. There are different types of standard fittings that the user can select from available standard libraries. Fitting is performed automatically, once the user selects the boundaries of each object manually.

The modelling of pipelines is summarized in three basic steps: (a) automated extraction of cylinders, (b) semantic classification of cylinders and (c) manual extraction and editing of pipes. Fitting is performed automatically for object extraction: therefore, it is not a separate step of the procedure. The average time per cylinder or pipe for each step is computed in our operating system in Table 9. The processing times for steps (a), (b) and (c) are calculated as following:

Table 9
Modelling time of each modelling task for each dataset and average time per object (min/object).

Modelling task	Typical facility Room 1	Typical facility Room 2	Water facility	Petrochemical plant	Average time (min)
	Time (min)				
Automated extraction of cylinders[a]	$1.5*10^{-3}$	$1.5*10^{-3}$	$7.1*10^{-3}$	$1.4*10^{-3}$	$3.3*10^{-3}$
Semantic classification of cylinders[b]	0.20	0.47	0.17	0.12	0.24
Manual extraction & editing of pipes[c]	0.69	2.37	2.43	1.22	1.68

[a] Per cylinder $*$ point.
[b] Per cylinder.
[c] Per pipe.

Table 10
Root Mean Square Error (RMSE) of the radius and coverage area (%) of automatically detected cylinders in each dataset and their average values.

Automatically detected cylinders in	RMSE of the cylinder radius (m)	Coverage area (%)
Typical facility - Room 1	$1.7*10^{-3}$	32.5
Typical facility - Room 2	$6.7*10^{-3}$	30.2
Water facility	$1.9*10^{-3}$	26.5
Petrochemical plant	$4.2*10^{-3}$	27.6
Average	$3.6*10^{-3}$	29.2

$$\text{time/cylinder. point} = \frac{\text{time for automated extraction of cylinders}}{\text{automatically detected cylinders}*\text{points in the dataset}} \quad (8)$$

$$\text{time/cylinder} = \frac{\text{time for semantic classification of cylinders}}{\text{automatically detected cylinders}} \quad (9)$$

$$\text{time/pipe} = \frac{\text{time for manual extraction\&editing of pipes}}{\text{total number of pipes}} \quad (10)$$

where the number of automatically detected cylinders and the total number of pipes is shown in Table 7 for each case study. The latter is the sum of automatically and manually detected pipes in each dataset. These normalizations are used to compare the modelling times for each case study, since the number of points and cylinders processed are different for each dataset.

The time for semantic classification was 0.24 min per cylinder on average. Manual extraction and editing of pipes was the most time-intensive step, since we needed 1.68 min per pipe on average to manually add missing pipes and edit the existing ones. The observations show that the manual effort to classify and extract pipes was 1.92 min per cylinder on average, which is the summation of two subsequent steps, (b) and (c). This is almost three times the time needed for automated extraction of cylinders by the software.

A variation of the time needed for automated extraction of cylinders between the water facility and the other datasets is observed. This discrepancy is attributed to the fact that the water facility is an outdoor facility, requiring the most processing time compared to the other datasets. Technically, outdoor scenes are inherently more occluded and incomplete exhibiting extreme variations in point density [43]. These effects are mitigated by the limited size and constrained shape of rooms. The two rooms of the typical industrial facility were processed at the same time in our operating system, for this reason the time required for automated extraction is the same as shown in Table 9. Manual modelling of the second room of this facility required the most modelling time. This is due to cluttered pipelines, which resulted in the largest Room Mean Square Error (RMSE) of the cylinder diameters, as shown in Table 10. This clutter is attributed to the reflective surface of pipelines. Manual extraction and editing of pipes in the water facility is another modelling time outlier. Highly occluded pipelines are the primary reason for this outlier, since they have the lowest average coverage (26.5%), compared to the other projects. The diameter of pipelines in this facility was significantly larger, since most pipes are used for sewage purposes. These observations show that manually detected pipes have larger average diameter (0.617 m) compared to automatically extracted cylinders (0.315 m) for the same dataset. This means that it is difficult for the software to identify cylinders with large diameters.

RMSE and coverage percentages for each extracted cylinder are calculated in the "*SmartSheet*", provided in EdgeWise. Table 10 summarizes their average values for all case studies. The results show that the first room of the typical facility has the lowest RMSE, meaning that the automatically extracted cylinders fit well the corresponding points of the cylinders. The average coverage area of cylinders in all case studies is around a quarter of the cylinder (29.20%), which is the reason that many cylinders are not automatically extracted.

The performance of the software is evaluated based on the two-following metrics, precision and recall [44],

$$\text{precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (11)$$

$$\text{recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (12)$$

where TP are the number of objects that are automatically detected as pipes and were correctly inspected as pipes.

FP are the number of objects that are detected as pipes, but we classified them as other cylindrical objects (for instance handrails, circular hollow steel sections to name a few).

FN are the number of objects that are pipes but were not automatically detected as pipes. Those pipes were manually extracted and added to the model.

The performance metrics obtained from our four sample datasets are given in Table 11. According to precision, out of all the automatically detected cylinders only an average of 47% in all case studies correspond to pipes, whereas the rest were other cylindrical objects. The average recall was 58.1%, meaning that only 58.1% of all pipes existing in a typical facility will be automatically detected. The results show that the water treatment facility, which is an outdoors facility, has the lowest recall, being 33.3%. The low performance metrics of this dataset, compared to the other ones, can be attributed to increased noise. The low precision of pipes in the first room of the typical facility (27.9%) is attributed to a larger number of FPs (roof tiles), which were wrongly detected as pipes.

The same metrics were measured for cylinders. The only difference in the metrics used is that precision is defined as the number of automatically detected cylinders out of all the detected cylinders, whereas recall is the number of automatically detected cylinders out of all other automatically detected non-cylindrical shapes. The recall of cylinders is high for all datasets except the petrochemical plant (45.7%), which is attributed to low scan completeness of this dataset and increased clutter. The average recall for the four datasets is 75.6% indicating the advantage of the software to extract this primitive shape. The precision of cylinders is also 15% higher compared to that of pipes, since the software is designed to detect cylindrical shapes. The lowest precision

Table 11
Average performance metrics of pipe and cylinder detection.

Dataset	Pipe detection metrics		Cylinder detection metrics	
	Recall (%)	Precision (%)	Recall (%)	Precision (%)
Typical facility Room 1	80.1	27.9	69.3	48.2
Typical facility Room 2	59.5	54.6	100.0	22.0
Water facility	33.3	36.4	87.3	86.4
Petrochemical plant	59.6	69.3	45.7	91.9
Average	58.1	47.0	75.6	62.1

(22%) is observed for the second room of the typical industrial facility, which is attributed to corrugated shapes in the roof that were incorrectly modelled as cylinders. The same trend (low precision of about 48%) is observed for the first room of the facility for the same reason.

3.6. Modelling of structural components

The most frequent structural elements that were identified above (CHSs, channels, I-beams) are modelled in the four case studies. The user selects the *I-Beam, Channel, Round Tubing* tools to manually extract the respective elements. The user can also create custom standards for shapes that do not exist on the standards list. The "*Pattern Extract*" tools extract groups or repeatable elements of the same object type. The extracted sections are then inspected for accuracy in the "*SmartSheet*".

The standards that were used for this evaluation were taken from the AISC manuals. We also used the "*Autofit*" tool to find the correct size of the specified section automatically. Precision and recall metrics were not used herein, since the procedure is manual.

3.7. Overall performance of state-of-the-art modelling software

Representative 3D models obtained from a room of a petrochemical plant are presented in Fig. 6. The laser scanned data was provided by AVEVA Group plc. The initial point cloud, the automated pipeline extraction output and the 3D model that was obtained after manual modelling of the most frequent pipeline elements, structural sections and electrical conduit are presented in Fig. 6. These 3D models are not the complete 3D models of the facility, but the subsets used for the evaluation purposes of this paper.

The purpose of Fig. 6 is not to show the complete 3D model of the facility, but rather an indicative example of the number of objects and the modelling time that took to model those. It is noteworthy that solid bars are not modelled separately in EdgeWise since they cannot be distinguished from circular hollow sections in a laser survey.

The 3D models can be exported to Revit, in order to obtain IFC models for interoperability purposes between different software packages. However, we observed that reducers, valves, flanges, angles and some channels (C3 and C4 according to the American Institute of Steel Construction standards - AISC) cannot be exported in Revit. Models containing straight elements with length < 4 mm cannot also be transferred to Revit.

The performance of state-of-the-art modelling software is summarized in Table 12. This Table shows that fitting of the most important object types has been solved by commercial software like EdgeWise, since known geometric shapes are automatically fitted to the selected point clusters.

Automated primitive shape detection of cylinders has partially been solved since the results showed 75% recall and 62% precision in EdgeWise. Non-cylindrical shapes are manually extracted, and classification of all object types has not been achieved.

Pipes, conduit and CHSs were also modelled manually in Revit to compare the man-hours needed for their shape extraction through this manual process. 30 objects were modelled in each category and their average modelling times were measured. The workflow of manual modelling in a software such as Revit entails three steps: (a) manual segmentation of the desired object in a point cloud visualization software such as CloudCompare, (b) export of the points in Autodesk Recap to obtain the appropriate format and then (c) modelling in Revit. Revit 2017 was used for this evaluation. The parameters of the cylinders (radius and length) are chosen based on the modeler's discretion.

3.8. Results

The time needed to model the above-mentioned object types is measured in the same operating system as stated above for pipeline, structural and electrical object types. The average modelling time per object for the most frequent object types is calculated. The manual modelling time of cylindrical objects is broken down to the two steps investigated above; shape extraction and semantic classification. Knowing the average number of objects of a specific type in a typical facility, we calculate the average modelling time for each object type and each modelling step where applicable. Fig. 7 shows the modelling time/object in minutes and Fig. 8 the estimated total man-hours for modelling of the same object types in a typical industrial facility in hours.

Fig. 7 shows that manual extraction of straight pipes in EdgeWise is the most time-intensive task compared to semantic classification for pipes that requires 1.68 min/straight pipe. Manual extraction of channels is also a laborious task compared to the manual extraction of all other object types, requiring 1.78 min/channel due to the complexity of their shape. Although some of the CHSs are automatically extracted, it is difficult to identify them manually, since they are usually pipe supports and handrails, which are significantly occluded. For instance, pipe supports are occluded due to pipelines that run on top of them. This is the reason for intensive modelling time (0.93 min/CHS). Semantic classification of cylinders is not a time-intensive step, requiring < 0.5 min/cylinder on average.

We choose a facility from the frequency-based case studies investigated above that has the median total number of objects out of the five facilities investigated. This is the petrochemical plant with 240,687 objects. Fig. 8 shows that pipes require the most modelling time on average (around 5200 labor hours) for this facility with 53,834 pipes. It is important to note that, although automated extraction of cylinders has been achieved by EdgeWise Plant/MEP, modelling of pipelines takes still substantial amount of time. The cylindrical shape is the most frequent geometric shape, thus the modelers' effort to distinguish electrical conduit, CHSs, handrails and other cylindrical objects from straight pipes is significant.

Although electrical conduit is the most prevalent object type in industrial plants (24.3% in a typical plant, Table 6), it takes less man-hours to model it compared to a straight pipe. This is attributed to the design of electrical conduit that places many cylinders closely to each other. This makes it easier for the modeler to identify them, thus the modelling time is reduced.

Flanges and elbows do not require substantial time (0.28 and 0.39 min/object respectively) as shown in Fig. 7, although the user manually adds them in the pipeline model. We observe that once the piping network is identified, the addition of fittings is a quick task that does not necessarily need to be automatically modelled. Angles require the least amount of time, being < 0.25 min/angle, which is attributed to their simple geometry compared to I-beams or channels.

The total labor hours for manual modelling of the petrochemical plant with 240,687 objects of the above categories are estimated to be 21 person-months. This finding is based on the following assumptions: (a) one trained modeler for all case studies, (b) the working hours are assumed to be 8 h/day, 5 days/week and (c) the operating system is as specified above. The same metric for cylinder extraction and classification is 17 person-months using EdgeWise as explained above. The

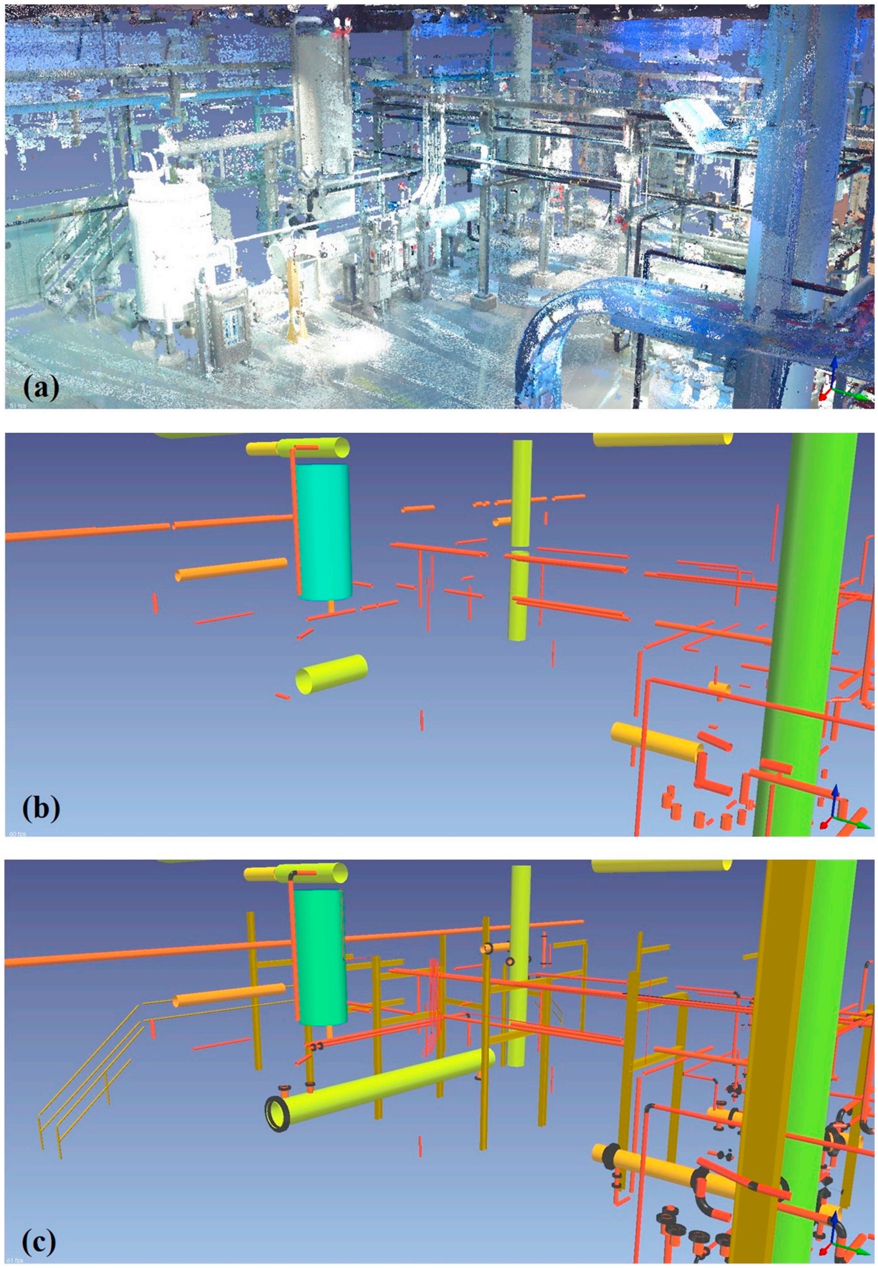

Fig. 6. (a) Input point cloud, (b) automated cylinder extraction in EdgeWise Plant/MEP and (c) 3D model after manual modelling of pipes, structural elements and electrical conduit for a room of a petrochemical plant.
(dataset provided by AVEVA Group Plc.)

Table 12
Performance of state-of-the-art software packages on each modelling step for the most important object types.

Industrial object type	Primitive shape extraction	Semantic labelling (classification)	Fitting
Straight pipe	Partially solved	Not solved	Solved
CHS[a]	Partially solved	Not solved	Solved
Channel	Not solved	Not solved	Solved
Conduit	Partially solved	Not solved	Solved
I-beam	Not solved	Not solved	Solved
Valve	Not solved	Not solved	Solved
Elbow	Not solved	Not solved	Solved
Flange	Not solved	Not solved	Solved
Angle	Not solved	Not solved	Solved

[a] Circular Hollow Section (CHS).

confidence intervals for the average manual modelling time of pipes, conduit and CHSs are calculated since the selection of parameters depends on the modeler's discretion. Pipes were manually modelled in 5.8 ± 1 min with 99% confidence level. Conduit and CHSs were modelled in 1.3 ± 0.75 and 3.6 ± 0.4 min respectively. This means that the modelling time does not change substantially for any of these object categories.

We observe that 64% of the man-hours needed for manual modelling of cylinders are saved by using the state-of-the-art software, EdgeWise, compared to conventional manual modelling platforms such as Revit. The results also show that 67% of manual modelling time is saved for pipe modelling. This case shows that 4836 labor hours are saved when modelling cylinders in EdgeWise. This is crucial especially for these facilities, since the time required to take decisions for maintenance and refurbishment is limited due to continuous production flow.

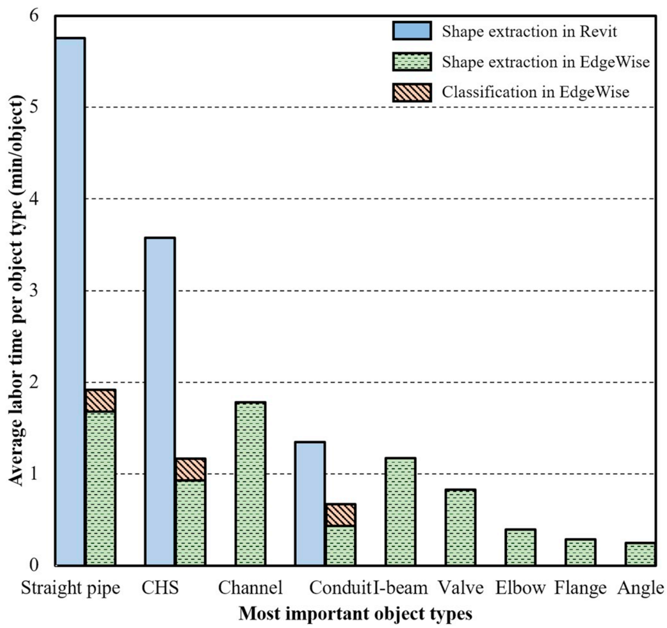

Fig. 7. Average modelling labor time per object (min/object) for the most important object types.

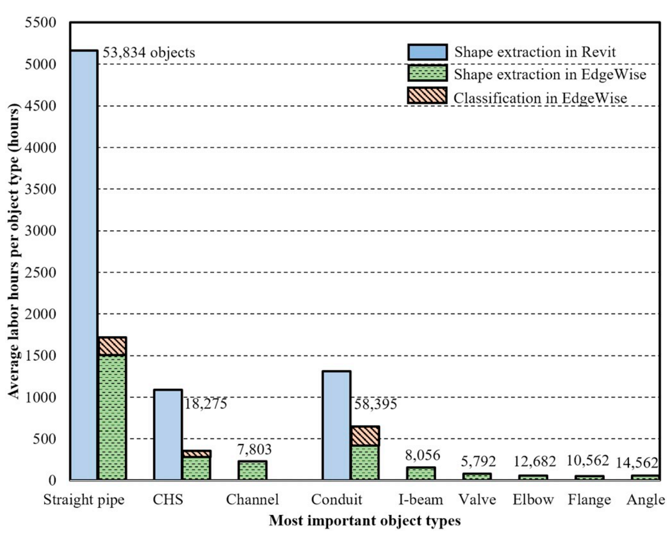

Fig. 8. Average modelling labor hours per object type for the most important objects of a sample facility with shown numbers of objects.

4. Conclusions

The ten most important object types in the three most frequent industrial categories (structural elements, piping network and electrical equipment) are ranked based on their frequency of appearance and modelling time. The results showed that cylindrical objects (straight pipes, electrical conduit and circular hollow sections) require 80% of the total modelling time of the ten most important object types in EdgeWise and represent 45.5% of the total number of objects in an industrial plant on average.

This paper marks the first study specifically aimed at identifying the most frequent and laborious to model industrial object types. The results of this paper show that current practice has achieved primitive shape extraction for straight pipes, elbows and conduit automatically. However, semantic labelling of each object type is not performed in the state-of-the-art modelling packages. Researchers have proposed methods for semantically labelling industrial objects based on prior knowledge, however these methods do not distinguish object classes that have the same primitive shape such as cylindrical objects (e.g., handrails, pipes, electrical conduit, circular hollow sections).

EdgeWise was selected compared to other state-of-the-art software, because it is the only commercially available tool that attempts to automatically extract cylinders from the point cloud of an industrial plant without significant user assistance. PointSense Plant has a similar functionality ("Walk the Run") as discussed above, however the user needs to identify the potential regions where the pipes are most likely to be located. This semi-automated approach guides the user through the run by suggesting insertion points for pipes and keeping the user in control of the modelling process. Therefore, EdgeWise was preferred since it has achieved greater level of automation for cylinder extraction. It has also substantially facilitated 3D modelling of industrial plants according to the findings discussed above. However, it has some limitations, which can be summarized as follows:

(a) The modeler should identify the structural elements manually or define the location of an object roughly in the point cloud to fit it.
(b) Detection of cylinders has only been partially solved, since cylinders are detected with 75% recall and 62% precision. The same metrics for pipes are 58% and 47% respectively.
(c) EdgeWise does not enrich the 3D geometric primitives with semantic labels and topological relationships. Engineers are required to manually implement the semantic labels of the components of the 3D model.
(d) Data inconsistency between different software platforms impedes modelers from exchanging data between different AI-BIM platforms. EdgeWise is not designed to provide a final output in an open and generic schema.

The contributions presented in this paper are therefore (1) the discovery of the most frequent object types in industrial facilities and their respective modelling time and (2) the measurement of the performance of state-of-the-art software and specifically EdgeWise. The latter uncovered (a) the substandard performance of this software in detecting cylinders, (b) the inability of this software to (i) further classify cylinders into conduits or pipes or CHSs and (ii) detect and further classify I-beams, channels elbows, flanges, valves and angles in spite of their high frequency in an industrial facility.

Direct implications of modelling the priority list of object types are assessed based on modelling time. The results of the evaluation of EdgeWise showed that semi-automatically modelling cylinders will reduce man-hours needed for modelling those by 64%. This can have a direct impact for industrial facility managers, since every hour of as-is modelling time is crucial for the operation of the plant in unprecedented circumstances (failures of critical objects, retrofitting operations and plant expansion).

Indirect implications of prioritizing object types are reductions of the modelling cost, since man-hours of modelers will be reduced. Although there is no way to calculate the exact cost of overestimated severity of industrial inspections and maintenance, it is reasonable to predict that maintenance of industrial plants will be substantially facilitated once AI-BIMs are easy to develop and the costs do not counteract the benefits of their creation. Poor maintenance of these assets does not always affect the asset's territory but also impacts nearby regions and puts lives of the public living close by at serious risk.

The presented research has room for improvement and some limitations of this study can direct future research. This study focuses on the industrial objects that are important to model, however methods on how to automatically model those were not investigated. Future work involves implementation of automated classification algorithms (e.g. machine learning) for the most important object types to minimize the modelling time. Application of these algorithms for hundreds of classes of different objects that have strong similarities (e.g., pipes, electrical conduit, CHs) is a very difficult multi-classification problem, that will be substantially benefited from the results of this exploratory research for the important objects to model in these complex environments. Overall, a training library of the object classes that are critical for industrial facility operations, frequent in industrial environments and laborious to model can be established to assist further research aimed at automated detection of these classes. Application of the findings of this paper will guide researchers on investigating methods for automatically modelling these objects.

Acknowledgements

We thank our colleague Bob Flint from BP International Centre for Business and Technology (ICBT), who provided insight and expertise that greatly assisted this research. We also express our gratitude to Otto Weiberth, Managing Director of ClearEdge^{3D}, Inc., who provided EdgeWise Suite for evaluation. The research leading to these results has received funding from the Engineering and Physical Sciences Research Council (EPSRC). AVEVA Group Plc. and BP International Centre for Business and Technology (ICBT) partially sponsor this research under grant agreements RG83104 and RG90532 respectively. We gratefully acknowledge the collaboration of all academic and industrial project-partners. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the institutes mentioned above.

References

[1] R. Volk, J. Stengel, F. Schultmann, Building Information Modeling (BIM) for existing buildings - literature review and future needs, Autom. Constr. (2014), https://doi.org/10.1016/j.autcon.2013.10.023.
[2] Cabinet Office, H., Government Construction Strategy, Construction, (2011) (https://doi.org/Vol 19).
[3] S. Tornincasa, F. Di Monaco, The future and the evolution of CAD, International Research Conference 'Trends in the Development of Machinery and Associated Technology', 18 2010.
[4] K. Kawashima, S. Kanai, H. Date, As-built modeling of piping system from terrestrial laser-scanned point clouds using normal-based region growing, J. Comput. Des. Eng. 1 (2014) 13–26, https://doi.org/10.7315/JCDE.2014.002.
[5] T. Rabbani, F.A. van den Heuvel, G. Vosselman, Segmentation of point clouds using smoothness constraint, International Society for Photogrammetry and Remote Sensing- Commission V Symposium 'Image Engineering and Vision Metrology', 36 2006, pp. 248–253.
[6] H. Son, C. Kim, C. Kim, Knowledge-based approach for 3D reconstruction of as-built industrial plant models from laser-scan data, International Symposium for Automation and Robotics in Construction, 756 2013, pp. 885–893.
[7] H. Veldhuis, G. Vosselman, The 3D reconstruction of straight and curved pipes using digital line photogrammetry, ISPRS J. Photogramm. Remote Sens. 53 (1998) 6–16, https://doi.org/10.1016/S0924-2716(97)00031-2.
[8] M. Fumarola, R. Poelman, Generating virtual environments of real world facilities: discussing four different approaches, Automation in Construction, 2011, pp. 263–269, , https://doi.org/10.1016/j.autcon.2010.08.004.
[9] J.-F. Hullo, G. Thibault, C. Boucheny, F. Dory, A. Mas, Multi-sensor as-built models of complex industrial architectures, Remote Sens. 7 (2015) 16339–16362, https://doi.org/10.3390/rs71215827.
[10] J.M. Douglas, Conceptual Design of Chemical Processes, McGraw-Hill New York, 1988, https://doi.org/10.1002/jctb.280460308.
[11] S. Petitjean, A survey of methods for recovering quadrics in triangle meshes, ACM Comput. Surv. 34 (2002) 211–262, https://doi.org/10.1145/508352.508354.
[12] British Institute of Facilities Management, BIM and FM: bridging the gap for success [WWW document], FM Leaders Forum Discuss. Pap, 2012 https://www.bifm.org.uk/bifm/knowledge/research/leadersforum.
[13] J. Edwards, A. Townsend, Buildings under Refurbishment and Retrofit [WWW Document], Chartered Institute of Building, 2011, http://www.carbonaction2050.com/sites/carbonaction2050.com/files/document-attachment/Buildings under Refurb and Retrofit.pdf , Accessed date: 17 June 2018.
[14] M. Baker, I. Stanley, Assessing and Modelling the Uncertainty in Fatigue Crack Growth in Structural Steels [WWW Document], Heal. Saf. Exec., 2008, http://www.hse.gov.uk/research/rrpdf/rr643.pdf , Accessed date: 17 June 2018.
[15] Z. Kala, Reliability analysis of the lateral torsional buckling resistance and the ultimate limit state of steel beams with random imperfections, J. Civ. Eng. Manag. 21 (2015) 902–911, https://doi.org/10.3846/13923730.2014.971130.
[16] E.A. Mechhoud, M. Rouainia, M. Rodriguez, A new tool for risk analysis and assessment in petrochemical plants, Alex. Eng. J. 55 (2016) 2919–2931, https://doi.org/10.1016/j.aej.2016.05.013.
[17] T.R. Moss, J.E. Strutt, Data sources for reliability design analysis, Proc. Inst. Mech. Eng. Part E J. Process Mech. Eng. 203–210 (207) (1993) 1989–1996, https://doi.org/10.1243/PIME_PROC_1993_207_203_02 13–19.
[18] E. Ruijters, M. Stoelinga, Fault tree analysis: a survey of the state-of-the-art in modeling, analysis and tools, Comput. Sci. Rev. (2015), https://doi.org/10.1016/j.cosrev.2015.03.001.
[19] G. Thompson, Improving Maintainability and Reliability Through Design, Professional Engineering Publishing Limited London and Bury St Edmunds, UK, 1999.
[20] A.A. Umar, Design for Safety Framework For Offshore Oil and Gas Platforms, (2010) (Ph.D. Thesis).
[21] US Chemical Safety and Investigation Board, Investigation Report Refinery Explosion and Fire BP Texas City, Investig. Rep. 1–341, 2007 (https://doi.org/REPORT No. 2005-04-I-TX).
[22] L. Landau, Piper Alpha: Lessons Learnt [WWW Document], Oil Gas, U.K., 2008https://oilandgasuk.co.uk/wp-content/uploads/2015/05/HS048.pdf , Accessed date: 17 June 2018.
[23] B. Singh, J.N. Britton, Offshore risk based corrosion integrity management a new methodology, Corros. National Association of Corrosion Engineers (NACE) International, 2001.
[24] G. Gourlis, I. Kovacic, Building information modelling for analysis of energy efficient industrial buildings - a case study, Renew. Sust. Energ. Rev. 68 (2017) 953–963, https://doi.org/10.1016/j.rser.2016.02.009.
[25] Autodesk, An Engineer's Guide to Capturing Energy Retrofit Opportunities [WWW Document], http://images.connect2communities.com/pdf/adeskwhitepaperaug2011.pdf, (2011) , Accessed date: 17 June 2018.
[26] D. Keeley, S. Turner, P. Harper, Management of the UK HSE failure rate and event data, J. Loss Prev. Process Ind. 24 (2011) 237–241, https://doi.org/10.1016/j.jlp.2010.09.002.
[27] M. Spain, P. Perona, Measuring and predicting object importance, Int. J. Comput. Vis. 91 (2011) 59–76, https://doi.org/10.1007/s11263-010-0376-0.
[28] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, A. Torralba, Scene parsing through ADE20K dataset, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5122–5130, , https://doi.org/10.1109/CVPR.2017.544.
[29] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (2015) 211–252, https://doi.org/10.1007/s11263-015-0816-y.
[30] Y. Xiang, R. Mottaghi, S. Savarese, Beyond PASCAL: a benchmark for 3D object detection in the wild, 2014 IEEE Winter Conference on Applications of Computer Vision, WACV 2014, 2014, pp. 75–82, , https://doi.org/10.1109/WACV.2014.6836101.
[31] FARO, FARO [WWW Document], https://www.faro.com/en-in/products/construction-bim-cim/faro-pointsense/features/, (2017) , Accessed date: 6 January 2018.
[32] M.F. Ahmed, C.T. Haas, R. Haas, Automatic detection of cylindrical objects in built facilities, J. Comput. Civ. Eng. 28 (2014) 04014009, , https://doi.org/10.1061/(ASCE)CP.1943-5487.0000329.
[33] A.K. Patil, P. Holi, S.K. Lee, Y.H. Chai, An adaptive approach for the reconstruction and modeling of as-built 3D pipelines from point clouds, Autom. Constr. 75 (2017) 65–78, https://doi.org/10.1016/j.autcon.2016.12.002.
[34] M.M. Sharif, M. Nahangi, C. Haas, J. West, Automated model-based finding of 3D objects in cluttered construction point cloud models, Comput. Civ. Infrastruct. Eng. 32 (2017) 893–908, https://doi.org/10.1111/mice.12306.
[35] Y. Perez-Gallardo, J.L.L. Cuadrado, Á.G. Crespo, C.G. de Jesús, GEODIM: a semantic model-based system for 3D recognition of industrial scenes, Intelligent Systems Reference Library, 2017, pp. 137–159, , https://doi.org/10.1007/978-3-319-51905-0_7.
[36] Everything 3D [WWW Document], http://www.aveva.com/en/solutions/product_finder/aveva_everything3d/, (2017) , Accessed date: 5 January 2018.
[37] R.A. Parisher, R.A. Rhea, Pipe drafting and design, Pipe Drafting and Design, 2012, https://doi.org/10.1016/B978-0-12-384700-3.00005-0.
[38] Engineering Toolbox [WWW Document], Pipe Lengths, https://www.engineeringtoolbox.com/pipe-length-d_744.html, (2004) , Accessed date: 15 June 2018.
[39] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C: The Art of Scientific Computing, 2nd ed., Cambridge University Press, 1992, https://doi.org/10.1063/1.2820230.
[40] H.M. Walker, De Moivre on the law of normal probability, A Source B. Math. 1985, pp. 75–84.
[41] D.M.W. Powers, Applications and explanations of Zipf's law, NeMLaP3/CoNLL '98 Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning, 1998, pp. 151–160, , https://doi.org/10.3115/1603899.1603924.
[42] ClearEdge, ClearEdge [WWW Document], http://www.screencast.com/users/ClearEdge3D, (2017) , Accessed date: 30 January 2018.
[43] T. Hackel, J.D. Wegner, K. Schindler, Contour detection in unstructured 3D point clouds, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1610–1618, , https://doi.org/10.1109/CVPR.2016.178.
[44] D.M.W. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation, J. Mach. Learn. Technol. 2 (2011) 37–63 (10.11.214.9232).
[45] F.H. Sanders, 3D laser scanning helps chevron revamp platform, Oil Gas J. 99 (2001) 92–98.
[46] R.J. Larsen, M.L. Marx, Introduction to mathematical statistics and its applications, fifth edition, (2012) 978-0-321-69394-5.