Automated detection of fault types in CCTV sewer surveys

Joshua Myrans, Richard Everson and Zoran Kapelan

ABSTRACT

Sewers must be regularly inspected to prioritise effective maintenance, which can be an expensive and time-consuming process. This paper presents a methodology to automatically identify the type of a detected fault using raw closed circuit television (CCTV) footage. The procedure calculates the GIST descriptor of a video frame containing a fault before applying a collection of random forest classifiers to identify the fault's type. Order oblivious filtering is used to further improve the methodology's performance on continuous footage. The technology, including various classifier architectures, has been validated and demonstrated on CCTV footage collected by Wessex Water. The methodology achieved a peak accuracy of 73% when applied to well-represented fault types, showing promise for future application in the water industry.

Key words | automated, fault analysis, GIST, random forest, sewers

Joshua Myrans (corresponding author)
Centre for Water Systems,
University of Exeter,
Harrison Building, North Park Road, Exeter,
Devon, UK
E-mail: *jm494@exeter.ac.uk*

Richard Everson
Department of Computer Science,
University of Exeter,
Harrison Building, North Park Road, Exeter,
Devon, UK

Zoran Kapelan
Delft University of Technology,
Faculty of Civil Engineering and Geosciences,
Department of Water Management,
Delft, The Netherlands

INTRODUCTION

Water companies all over the world are tasked with the effective maintenance of extensive sewer networks. This asset management is normally performed based on information gathered via regular CCTV (closed circuit television) surveys, which record and detail the otherwise inaccessible interior of the pipes. The footage is captured using a camera, which can be driven through the system, using a remotely controlled PIG (pipe inspection gadget) or pushed through the pipes using a system of push rods. Due to the additional control provided by the PIG, this technique usually acquires better quality footage; however, both techniques are sufficient for a human technician to perform an effective analysis of the pipe. Any footage collected in the UK is annotated according to the *Manual of sewer condition classification* (WRc 2013), highlighting faults and labelling them according to strict definitions. This analysis is often performed offline, requiring a technician to re-watch the entire length of footage annotating each fault. Alternatively, technicians using a setup with a real-time video feed can annotate the footage online, labelling faults as they are recorded.

Current surveying practices can provide detailed information for pipes, although the process of doing so is time-consuming and expensive. Offline surveys are the slowest, requiring the entire duration (usually hours) of footage to be re-watched, to get a comprehensive analysis. Online surveys are normally faster, although the survey is often interrupted, leaving the camera stationary while the operator records a fault. If the speed of surveys could be increased, more surveys could be performed, giving water companies better coverage of their network. It is not uncommon that water companies must pay to access their assets, especially if this requires the disruption of local infrastructure. These fines often depend on the duration of the disruption, as such, improving the speed of surveys could also reduce their cost. Finally, as surveys are performed by humans they can be subjective and are prone to human error, as discussed by van der Steen *et al.* (2014). This lack of uniformity is often exacerbated when water companies rely on multiple external contractors for their sewer surveys, as all have different working practices.

The field of automated fault analysis contains multiple effective methodologies, with machine learning techniques

consistently demonstrating success. Some of the earliest work was performed by Duran *et al.* (2007), who retrofitted existing CCTV cameras with a laser profiler. This device recorded precise interior dimensions of the pipe, which were passed to an artificial neural network (ANN) for classification. Around the same time, Sinha & Fieguth (2006) also applied ANNs and fuzzy logic to detect cracks in sewer pipes. A collection of key parameters, such as light intensity, shape and size, were extracted from individual frames. These measurements were transformed using fuzzy logic, incorporating expert knowledge, before the ANN determined the presence of a crack.

Useful in many engineering applications, the field of autonomous crack detection continued to develop. Jahanshahi & Marsi (2012) utilised a sequence of 2D images to construct the geometry of a pipe's interior, producing a 3D scene with which to identify a crack. This identification was performed using both a support vector machine (SVM) and ANN, and could determine key information about a crack, including length and depth. Halfawy & Hengmeechai (2013) used a collection of image processing techniques to identify cracks in sewer pipes, identifying edges in a frame using the Sobel method, before applying specialised filters to highlight cracks. Similarly, Khalifa *et al.* (2014) applied a Canny edge detector, inside a Markovian framework to detect cracks using temporal information. Most recently, Chen *et al.* (2017), produced a methodology for detecting cracks in metallic pipes. Using local binary patterns, a SVM and Bayesian decision theory, Chen *et al.* (2017) achieved an 85% hit rate, capable of distinguishing between cracks, scratches and welds.

Looking at more general fault detection, Guo *et al.* (2009a) identified anomalies in sewer pipes, using a process of frame differencing. This technique looked for sudden changes between frames of CCTV footage, indicating the presence of a fault in the pipe. Guo *et al.* (2009b) continued this work matching the scale invariant feature transform (SIFT) of neighbouring frames. Like the frame differencing technique, if neighbouring SIFT features are sufficiently dissimilar they are identified as a potential fault. Halfawy & Hengmeechai (2014a) developed a novel solution to fault detection, using the motion of the camera to detect faults. Given the pan and tilt capabilities of a PIG, the optical flow was calculated to identify the motion of the camera.

As the operator moved the camera to track faults within a pipe, the methodology would identify this abnormal motion and flag the footage as containing a fault. Halfawy & Hengmeechai (2014b) continued to detect faults using histogram of ordered gradients (HOG) features and SVM. Individual frames were split into grids of cells and the HOG features of each cell calculated, before being classified using the SVM. Demonstrated on tree root intrusions, the methodology proved to be effective, achieving a detection accuracy of 86%. Most recently, Hawari *et al.* (2018) developed an ensemble of techniques to detect cracks, deposits, ovality and displaced joints. Applied individually, each of these techniques showed promise for their respective fault types, achieving true positive rates (TPR) ranging from 53% to 74%.

The above methods approach the problem of fault detection by developing individual methods that work for specific sewer faults only. This way, once the fault is detected, its type is automatically known. The downside of this, however, is that this approach requires development of a whole suite of different detection methods which normally results in the development of a small number of methods addressing only several of the most frequent types of faults. As a consequence, some less common fault types may go undetected and could ultimately result in a sewer blockage or collapse leading to a pollution or flooding incident. Myrans *et al.* (2018) decided to take an alternative, fundamentally different approach and have developed a more general methodology that is capable of detecting faults of any type from raw CCW footage. Building upon the success of this detection methodology, this paper presents a fault classification methodology capable of categorising detected faults according to the *Manual of sewer condition classification* (WRc 2013). Categorisation of faults is performed using a combination of image processing and machine learning techniques, providing a prediction of each fault's presence within a frame.

METHODOLOGY

Problem definition and overview

The fault type identification methodology applies an ensemble of image processing and machine learning techniques to

CCTV frames that were detected to contain faults. This methodology is data-driven in the sense that all image characteristics and classifier parameters are learned from the labelled database of frames used to train the methodology. As such, it is assumed that the database contains a sufficient number of relevant examples of each required fault type. If the technology were to be applied in practice, the methodology's database of frames could be regularly updated to include images of all labelled frames from recent surveys. This constant feed of new information should continue to improve the effectiveness of the methodology, providing more examples of each fault. In addition, the training process requires no human interaction, and can be performed periodically overnight, minimising the impact of the lengthy (<30 minutes for this case study) training process.

When applied to continuous sequences of frames, each frame can be processed in turn, identifying the type of a present fault, as illustrated in Figure 1. The methodology's structure can be intuitively broken down into three stages: 'Pre-processing', 'Feature extraction' and 'Classification'.

Data pre-processing stage

This stage aims to reduce the complexity of the later classification problem, eliminating information unnecessary for the identification of faults. Given an RGB frame, known to contain a fault, it is converted to greyscale, and re-sized to a lower resolution (128×128 pixels). In combination, these steps reduce the number of values required to represent a standard CCTV image (512×512 pixels) to 1/48th of the original count. These steps were possible as experimentation showed colour to have little impact on the methodology's performance, likely due to the large variation in illumination that is often found in standard CCTV footage. Similarly, resolution can be reduced, as higher resolution images showed negligible improvement in performance, while dramatically increasing the methodology's running time.

As the methodology only distinguishes between fault types, a separate methodology is required to automatically detect the presence of a fault. Any fault detection methodology could be applied to achieve this; however, this methodology was especially designed to integrate with the

methodology developed by Myrans *et al.* (2018). As both methodologies perform the same 'Pre-processing' and 'Feature extraction' stages, each would only need to be performed once, saving on computational time.

Feature extraction stage

The 'Feature extraction' stage aims to further reduce the complexity of the later classification problem, converting the processed frame to a GIST descriptor (Oliva & Torralba 2001). A GIST descriptor describes the contents of the image using a series of Gabor wavelets at several orientations and scales. To calculate a standard GIST descriptor, a greyscale image is convolved here with a series of Gabor filters (Bovik *et al.* 1990), arranged at four scales and eight orientations (see Figure 2). The resulting 32 feature maps characterise the contents of the original image. To form the final descriptor, each of the 32 feature maps is overlaid with a 4×4 grid of square cells (see Figure 2). The contents of each cell in each map is finally totalled to form the final 512 value feature descriptor. Experimentation showed that increasing the number of scales, orientations or cells had minimal impact on the methodology's performance. Note that by using the GIST descriptor the 128×128 pixel greyscale images (generated during the pre-processing stage) are reduced from 1.6×10^4 pixel values to 512 numerical values only.

Other common feature descriptors, including HOG (Dalal & Triggs 2005) and SIFT (Lindeberg 2012) exist. These concentrate on specific image details, yielding a much higher dimensional image descriptor. However, preliminary experiments showed that HOG and SIFT descriptors, at best, performed similarly to GIST, while taking much longer to process. For this reason, HOG and SIFT descriptors were not used here.

Classification stage

The aim of this stage is to classify a frame's processed feature descriptor, according to the fault types specified by the *Manual of sewer condition classification* (WRc 2013). The most effective classification technique ('1 vs all') is discussed here. Details of other tested techniques can be found in the section 'Alternative classification methods'.
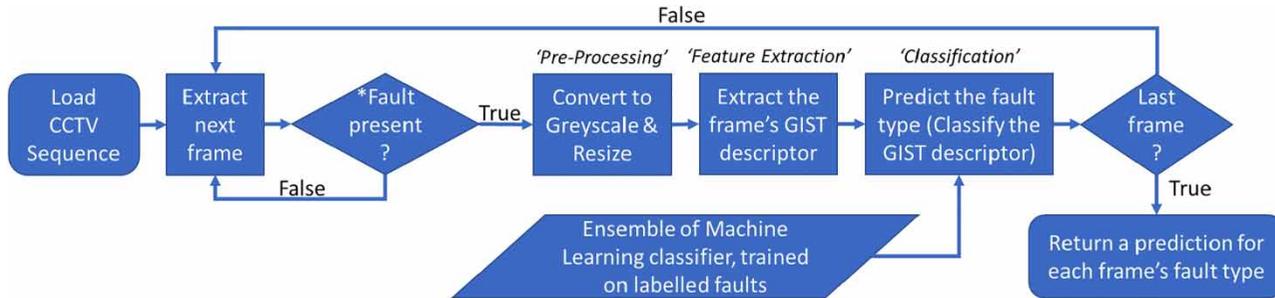
Figure 1 | Flow chart defining the structure of the fault identification methodology, when applied to continuous CCTV footage. *As this methodology does not predict the presence of a fault, another methodology should be used for this stage (see the discussion in the section below).

The '1 vs all' technique classifies a frame's contents by predicting the probability of each fault type's presence against all others, selecting the most likely category of fault. This is achieved using a collection of random forest (RF) classifiers (Breiman 2001), one for each fault type. Each RF then attempts to identify its given type from all other faults, returning a predicted probability of that fault type's presence. Once a predicted probability for each fault type is calculated, the faults are ranked from most to least likely. Finally, the highest ranked fault type is assigned to the frame, completing the classification. This simple approach breaks the multi-class classification into many simpler binary classifications, improving the effectiveness of the classifiers, which often perform better on these simpler tasks (Essid et al. 2006).

A RF is an ensemble classifier, which utilises a collection of decision trees to predict the class of a given digital object (Breiman 2001). Each tree predicts the type of a given fault as one of the categories in the training data. Averaging (voting) over the predictions made by all trees in the ensemble obtains a mean prediction which has been found to generalise well to unseen data. As with all supervised machine learning techniques, the RF classifiers must be pre-trained on a dataset of labelled frames, as discussed in the section 'Problem definition and overview'. This training is only required once and must be performed before the methodology is applied in practice. Due to its previous success (Myrans et al. 2018), the extremely randomised trees (extra trees) algorithm (Geurts et al. 2006) is used for this training process. In the extra trees' paradigm, a RF is calibrated, 'growing' each internal decision tree by randomly selecting a single value from the GIST feature descriptor to split the training dataset. In a traditional RF, the split is performed based on the maximum information gain, splitting the training dataset as evenly as possible (Breiman 2001). Conversely, the 'extra trees' algorithm selects a random threshold to perform the split, from a small number of randomly chosen features often resulting in unbalanced branches. This feature often leads to larger trees able to perform accurate classifications and generalise well to
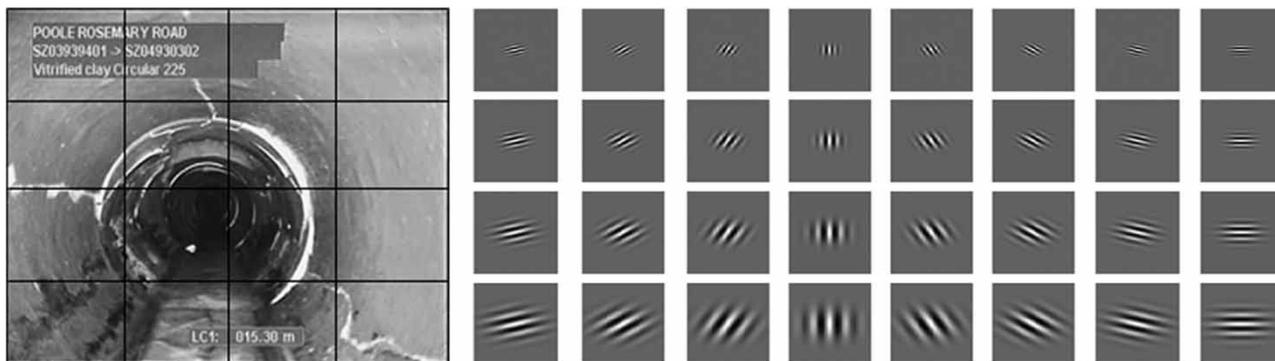

Figure 2 | (Left) Greyscale frame, overlaid with the 4×4 grid of square cells. (Right) Gabor filters at four scales (S) and eight orientations (θ).

unseen data (Geurts *et al.* 2006). Geurts *et al.* (2006) show that this calibration strategy results in a RF that classifies accurately and generalises well to new data. Once an entire forest is trained, unseen frames can be classified: GIST descriptors are processed by every tree in the forest, after which, the forest votes on a frame's class. The voting proportions can be interpreted in order to generate an estimated probability of class membership. These probabilities are those ranked in the '1 vs all' architecture to find the most probable fault type.

Alternative classification methods

The case study of this paper aims to compare three other classification techniques to the '1 vs all' methodology described above. These include:

- 'Single' RF: As random forests can perform multi-class classifications (Breiman 2001), a single RF was used as a benchmark for the other classification approaches. This requires very little internal adjustment to the RF's structure, instead of voting over two classes, decision trees vote over the 13 labelled fault categories. This implementation has the advantage of being the simplest approach, requiring (by far) the least time to process.

- 'Pairwise' classification: Pairwise classification trains a RF classifier to compare every fault type against every other fault type, resulting in $n^2 - n$ classifiers, where n is the number of fault categories. As each classifier is only making a binary decision between two fault types each class has $n - 1$ *associated predictions*. These predicted probabilities are then summed to give a score for each fault type, with every fault's score being ranked against the rest. Much like '1 vs all' classification, a frame is then assigned the type of the largest score.

- 'Weighted pairwise' classification: Weighted pairwise classification follows the same procedure as pairwise classification; however, each classifier's predicted probability is weighted before being summed to give a score (Hüllermeier & Vanderlooy 2010). By doing so it is hoped that bias within the dataset can be negated, adding weight to under-represented fault types and enabling a clearer separation between faults with similar

appearances. These weights are learnt using the CMA-ES evolutionary algorithm (Hansen *et al.* 2003). This algorithm runs during the training process, identifying the weights to be applied to unseen data. The workflow of this technique is presented in Figure 3.

CASE STUDY

Data

This case study uses data from real CCTV footage collected by UK water company, Wessex Water. The surveys cover over 30 km of pipe ranging from 150 to 1,500 mm in diameter and cover a variety of pipe shapes (circular, egg, horseshoe) and materials (vitrified clay, PVC and brick). From these surveys a selection of 2,260 faults were extracted and labelled according to the survey annotations. According to Wessex Water, these distributions of fault type, pipe material and shape are a good representation of their network. As the current methodology currently aims to identify singular fault types, all images which contained multiple fault types were combined to form a new 'multiple' category of fault, resulting in a final dataset of faults distributed as shown in Table 1.

It should be noted that all the work presented has been developed in Python, using the Anaconda distribution. The core classification techniques (random forests) have been implemented from the scikit-learn packages (Pedregosa *et al.* 2011). The classification architectures ('1 vs all', 'pairwise' and 'weighted pairwise') have been developed from scratch; however, suitable alternatives can also be found on scikit-learn.

Results and discussion: still images

This case demonstrates and compares the performance of four ensembles of machine learning classifiers when applied to the sewer fault type identification problem. The four techniques are: the 'single' (Myrans *et al.* 2018), '1 vs all', 'pairwise' and 'weighted pairwise' classifiers, as described in the sections 'Classification stage' and 'Alternative classification methods'.
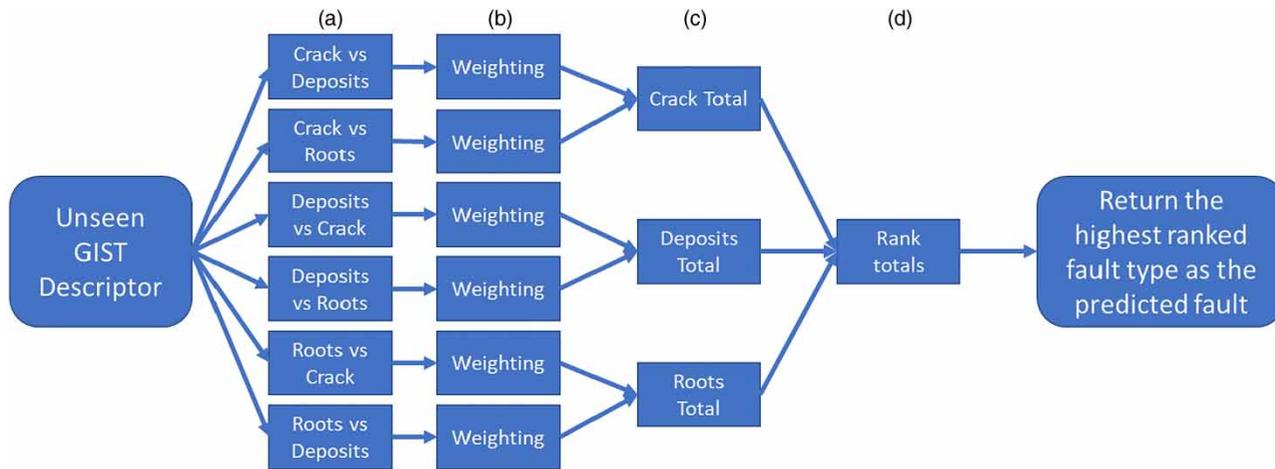
Figure 3 | Flow chart demonstrating the application of weighted pairwise classification to determine between 'cracks', 'deposits' and 'roots' fault types. (a) Application of each pairwise classifier, (b) multiplying each classifier's predicted probability by its associated weight, (c) summing the weighted predictions for each fault type and (d) ranking the weighted prediction total for fault types.

To objectively compare the four techniques, this case compares the raw accuracy defined as the percentage of frames correctly classified by each technique. However, this measure alone gives only a naive understanding of each technique's performance. Because of this, the confusion rate matrix for each technique is examined too, to

Table 1 | Distribution of fault types present in the Wessex Water dataset

Fault type	Subtype	Percentage (%)
Joint	Displaced, open	31.5
Deposits	Attached, settled	15.7
Multiple	–	12.7
Crack	Longitudinal, circumferential, multiple, spiral	10.0
Surface	–	9.9
Roots	Fine, tap, mass	8.2
Infiltration	Running, gushing	4.8
Obstacles	Intruding junctions, masonry, protrusion	2.8
Other	Vermin, lining	1.4
Broken/Collapsed	–	1.3
Hole	–	0.9
Brickwork	Missing mortar, Displaced bricks, Missing bricks	0.5
Deformation	–	0.3

Each fault may be further divided into subtypes as described, but only the major fault types were used for classification.

highlight the strengths and weaknesses of each approach. Finally, as each classification technique also ranks fault types from most to least likely in each frame, the accuracy over the most likely two, three and five predictions are considered. This last measure could be useful in the development of a decision support tool, working alongside a technician to offer a choice of two, three or five most likely faults when the methodology struggles to identify a single fault type.

To make best use of the available dataset, 25-fold cross validation was used to separate the frames into training and testing sets (Kohavi 1995). Cross validation splits the randomly shuffled dataset into 25 equally sized groups (i.e., folds). Each of the 25 folds is in turn set aside to form the test dataset, while the remaining 24 folds are used to train the classifier. The generalisation accuracy of the method is then estimated by averaging the accuracy over the 25 validation sets.

Once applied to the entire dataset of 2,260 frames and taking the technician's labels as the ground truth, the 'single', '1 vs all' and 'pairwise' techniques performed well, achieving accuracies of 62.5%, 63.3% and 62.3%, respectively. These accuracies indicate that these automated techniques are sufficiently accurate to aid surveyors in identifying fault types. On the other hand, the 'Weighted pairwise' classification achieved a much lower accuracy of 17.7%, struggling to make accurate predictions. When

examining the misclassifications in more detail it is clear that this technique's predictions have a strong bias to the 'multiple' fault type misclassifying over 70% of faults as 'multiple'.

In terms of confusion matrices, all classification techniques performed similarly. All techniques struggled to identify 'multiple', 'obstruction', 'brickwork', 'hole', 'other' and 'broken' fault types achieving TPR of less than 50% for each fault type (Table 2). In the case of the 'multiple' fault type, this outcome is likely due to the large variety of fault combination that the category can include, given this experiment covers 12 distinct categories of fault. The remaining fault categories ('obstruction', 'brickwork', 'hole', 'other' and 'broken') are likely misclassified due to their poor representation in the dataset, each having less than 100 examples making up less than 5% of the dataset. In all the above cases, training on a larger number of examples of these faults would likely improve the predictions. By doing so, all of the fault categories would be better representative of the wide variance of each fault's appearance.

Given that each classification (i.e., fault type identification) technique ranks faults in order of likeliness, the top ranked faults were explored. As before, the 'single', '1 vs all' and 'pairwise' classifiers, performed similarly well achieving a highest accuracy of 86% when considering the most likely three out of 13 fault categories. This high accuracy lends the automated methodology to a decision support role, offering a shortlist of fault types for frames the methodology is less sure of.

Finally, as experiments have shown that methodology struggles to classify poorly represented fault types, the methodology was applied to a reduced dataset. This dataset includes frames containing only single fault types with at least 100 observations. This lead to a reduction of the original dataset to 1,816 frames containing the 'infiltration', 'joint', 'deposits', 'roots', 'crack' and 'surface' fault types only. As can be seen from Table 3, eliminating under-represented faults significantly improved the accuracy of the techniques, with '1 vs all' classification achieving the highest accuracy of 74.3%, closely followed by 'single' at 73.4% and 'pairwise' at 73%. As in the case of the full dataset, 'weighted pairwise' classification performed poorly with an accuracy of just 32.5% and a high bias towards the most common

Table 2 | Confusion rate matrix for the '1 vs all' classifier

'1 vs All' classifier (63.3%)		Predicted class												
		Surface	Multiple	Deposits	Joint	Infiltration	Obstruction	Roots	Brickwork	Crack	Hole	Other	Deformation	Broken
Correct class	Surface	**57.8**	3.6	8.4	24.4	0.9	0.0	1.3	0.0	1.8	0.9	0.9	0.0	0.0
	Multiple	6.6	**30.2**	14.6	31.9	0.7	2.8	4.9	0.3	7.3	0.3	0.0	0.0	0.3
	Deposits	3.7	7.6	**69.4**	16.9	0.3	0.3	0.6	0.3	0.3	0.6	0.3	0.0	0.0
	Joint	2.0	4.5	3.1	**86.1**	0.8	0.0	0.1	0.0	3.1	0.0	0.1	0.0	0.1
	Infiltration	0.0	3.6	0.9	27.0	**65.8**	0.0	0.9	0.0	1.8	0.0	0.0	0.0	0.0
	Obstruction	1.6	21.3	19.7	8.2	4.9	**34.4**	3.3	0.0	4.9	1.6	0.0	0.0	0.0
	Roots	2.7	8.6	8.6	22.0	2.2	0.0	**54.3**	0.0	1.6	0.0	0.0	0.0	0.0
	Brickwork	0.0	18.2	63.6	0.0	0.0	18.2	0.0	**0.0**	0.0	0.0	0.0	0.0	0.0
	Crack	2.2	8.0	3.5	23.0	1.3	0.0	0.9	0.4	**60.2**	0.0	0.0	0.0	0.4
	Hole	10.0	35.0	5.0	15.0	0.0	5.0	5.0	0.0	5.0	**20.0**	0.0	0.0	0.0
	Other	16.1	16.1	9.7	16.1	3.2	0.0	0.0	0.0	9.7	3.2	**25.8**	0.0	0.0
	Deformation	0.0	0.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**80.0**	0.0
	Broken	3.6	28.6	21.4	21.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**25.0**

Bold values identify the True Positive Rates for each fault type.

Table 3 | Accuracy when considering the top ranked faults

	Accuracy (%)			
	Top 1	Top 2	Top 3	Top 4
'Single'	62.5	76.5	85.0	93.3
'1 vs all'	63.3	77.4	85.4	93.8
'Pairwise'	62.4	77.4	86.0	94.8
'Weighted pairwise'	17.7	56.2	76.4	91.3

'joint' category. Examining the confusion matrix of '1 vs all' classification for this reduced dataset (see Table 4), it is clear that other techniques also suffer with this bias towards the 'joint' class, albeit to a lesser extent with most non-joints faults being commonly misclassified as 'joints'. In order to eliminate this bias, the methodology could be trained on a uniform number of each fault type. Preliminary experiments show that this did reduce the bias, although providing minimal improvement in the technique's accuracy, as the misclassifications were instead distributed across all classes. By extension, alternative, more complex sampling strategies could be implemented. Samples could be chosen proportional to the number of examples, i.e., oversampling minority classes and undersampling majority classes. On the other hand, as the dataset's fault types are distributed similarly to those found in the wild (Table 1), it could be argued that this bias towards the most common fault type is desirable, being representative of faults in the field.

Results and discussion: video

In order to demonstrate the fault type identification technique's application on continuous real-life CCTV footage, the 'single' and 'pair' classification approaches were applied to a segment of unseen sewer survey. This case applies the best performing classification technique '1 vs all', trained on the full dataset of 2,260 frames to a 3-minute sewer survey containing 'roots', 'joint' and 'deposit' faults, each of which spanned multiple frames.

As the methodology is being applied to contiguous frames, a little extra information can be gained from neighbouring frames, when performing the fault type classification. For example, if the previous frames contain a 'root' fault, it is likely the next frame will also contain a 'root' fault. To incorporate this idiom into the methodology, the sequence of predictions is smoothed, using order oblivious filtering (Yan *et al.* 2012). This technique votes on the fault's type over the faults 50 neighbouring frames (25 frames each side of the frame in question). This simple smoothing technique has been selected over its alternatives, due to its previous success when applied to the fault detection problem (Myrans *et al.* 2018).

Over the entire duration of footage, the technique achieved an accuracy of 66.3% on frames labelled as containing a fault, in line with the accuracies seen in the section 'Results and discussion: Still images'. However, six out of seven blocks of faulty frames types were correctly identified, a breakdown of which can be seen in Figure 4.

Examining Figure 4 in more detail, faults 2, 4, 5, 6 and 7 are clearly identified, with over 90% of the fault's duration being correctly labelled, where a fault's duration is defined as the number of consecutive video frames the fault appears in. Fault 1 has been less convincingly classified, with only 33% of its duration being identified. This is likely due to the low severity of this 'joint' fault, making it hard to distinguish between a normal and displaced joint. Finally,

Table 4 | Confusion matrix for the '1 vs all' classification technique, when applied to the small well-represented dataset

		Predicted class					
		Infiltration	Joint	Deposits	Roots	Crack	Surface
Correct class	Infiltration	**0.5**	29.7	0.9	0.9	1.8	0.9
	Joint	0.7	**89.7**	4.4	0.4	2.9	1.8
	Deposits	0.8	21.3	**71.3**	0.8	0.8	4.8
	Roots	1.1	24.2	9.1	**60.2**	2.2	3.2
	Crack	1.3	27.9	3.5	0.9	**62.8**	3.5
	Surface	1.3	28.0	7.1	1.8	4.0	**57.8**

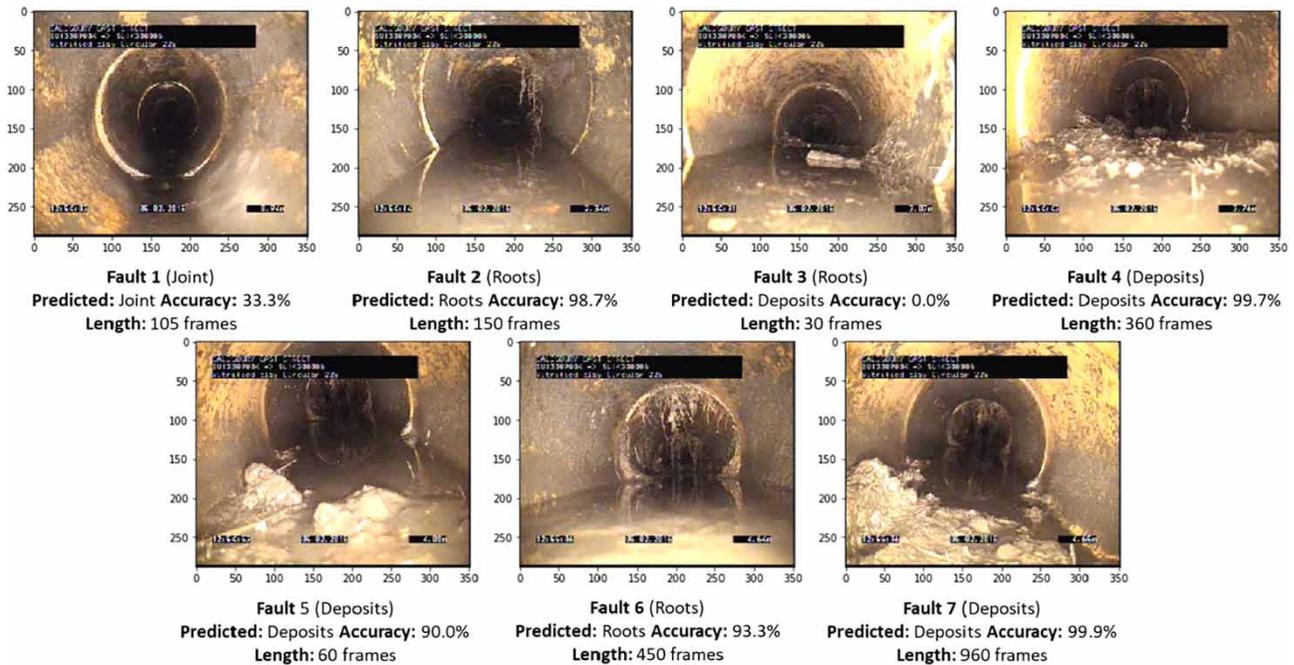Bold values identify the True Positive Rates for each fault type.

Figure 4 | The seven faults present in the continuous CCTV segment. Labelled with the true fault type, the predicted fault type, the duration of the fault, and the percentage of fault's associated frames correctly identified.

fault 3 has been completely missed, with 0% of its duration being correctly identified as a 'roots' fault. It was instead classified as a 'deposit' fault. Factors that could have led to this misclassification could include its short duration, discrete nature, and the arguable presence of multiple faults. As it is common for faults in sewers to appear in clusters, the presence of multiple faults is a topic of importance and requires further investigation. Suggested strategies for overcoming these issues include the implementation of a multi-labelling strategy. This would label a frame with all faults that lie above a given prediction threshold. To check the filtering was not negatively impacting the detection of faults 1 and 3, the experiment was re-run without filtering, but this achieved inferior results.

CONCLUSION

The work presented in the paper demonstrates a novel methodology for automatically identifying the type of a detected fault in sewer surveys. The proposed methodology calculates GIST feature descriptors for faulty frames and applies an RF machine learning classifier to analyse the frame's contents, identifying the fault type. The methodology builds on the Myrans *et al.* (2018) detection methodology that identifies the presence of a general fault in the image. The fault type identification methodology was tested, validated and demonstrated on both still images and continuous footage obtained from a UK water company. Based on the results obtained the following observations can be made:

1. The proposed fault type identification methodology is able to automatically identify a wide range of fault types in real-world sewer surveys with a relatively high accuracy. The methodology appears to be robust and reliable, working on both still CCTV images and video sequences.
2. The '1 vs all' classification architecture proved to be the most effective achieving a peak classification accuracy of 74% on well represented fault types. The 'single' and 'pairwise' architectures were also reasonably effective.
3. Order oblivious filtering assists the application of the methodology to continuous video footage, incorporating information from neighbouring frames to improve accuracy.
4. Particular care should be taken when selecting the calibration dataset as all required fault types should be well

represented (at least 100 examples). However, if this is not possible, alternative approaches can be considered (classifying only a subset of fault types or considering the most likely two, three or four fault types).

Future work will focus on the development of this technology, applying the fault type identification methodology to a larger and more comprehensive dataset. In addition, sampling strategies will be investigated and tested in an attempt to reduce the bias in predictions towards majority classes. The use of multi-label classification is also being tested, attempting to classify multiple faults present in a given frame. Finally, the methodology will be extended to identify other features within a sewer pipe, including connections and manholes. This idea could be further extended to identify the presence of illicit connections (Panasiuk *et al.* 2015) by overlaying a schematic of the sewer network with the tool's results.

As this work was developed to integrate with a previous fault detection methodology (Myrans *et al.* 2018), the next steps will be to combine the methodologies into a single decision support tool to assist technicians in the field. This decision support tool will be designed to assist the engineer and could investigate the addition of extra technologies. These could include the identification of the location of a fault within a frame using the RF's intuitive structure (Myrans *et al.* 2018) or incorporating additional information about the pipe/topology to improve decision-making. Overall, this fault type identification methodology, alongside other techniques, will work together to speed up and assist the surveying of sewer pipes.

ACKNOWLEDGEMENTS

REFERENCES

Bovik, A. C., Marianna, C. & Wilson, S. G. 1990 Multichannel texture analysis using localized dpatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12** (1), 55–73.

Breiman, L. 2001 Random forests. *Machine Learning* **45** (1), 5–32.

Chen, F., Jahanshahi, M. & Joffe, C. 2017 A texture based methodology using Bayesian data fusion for autonomous crack detection on metalic surfaces. *Computer-Aided Civil and Infrastructure Engineering* **32** (4), 271–287.

Dalal, N. & Triggs, B. 2005 Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition* **1**, 886–893.

Duran, O., Althoefer, K. & Seneviratne, L. 2007 Automated pipe defect detection and categorization using camera/laser-based profiler and artificial neural network. *IEEE Transactions on Automation Science and Engineering* **4** (1), 118–126.

Essid, S., Richard, G. & David, B. 2006 Musical instrument recognition by pairwise classification strategies. *IEEE Transactions on Audio, Speech and Language Processing* **14** (4), 1401–1412.

Geurts, P., Ernst, D. & Wehenkel, L. 2006 Extremely randomised trees. *Machine Learning* **63** (1), 3–42.

Guo, W., Soibelman, L. & Garrett, J. 2009a Automated detection for sewer pipeline inspection and condition assessment. *Automation in Construction* **18** (5), 587–596.

Guo, W., Soibelman, L. & Garret, J. H. 2009b Automated defect detection in urban wastewater pipes using invariant features found in video images. In: *Construction Research Congress 2009: Building A Sustainable Future*, Seattle, WA, USA.

Halfawy, M. & Hengmeechai, J. 2013 Afficient algorithm for crack detection in sewer images from closed-circuit television inspections. *Journal of Infrastructure Systems* **20** (2), 04013014.

Halfawy, M. & Hengmeechai, J. 2014a Integrated vision-based system for automated defect detection in sewer closed circuit television inspection videos. *Journal of Computing in Civil Engineering* **29** (1), 04014024.

Halfawy, M. & Hengmeechai, J. 2014b Automated defect detection in sewer closed circuit television images using histograms of oriented gradients and support vector machine. *Automation in Construction* **38**, 1–13.

Hansen, N., Müller, S. D. & Koumoutsakos, P. 2003 Reducing the complexity of the derandomized evolution strategy with covariance martrix asaptation (CMA-ES). *Evolutionary Computation* **11** (1), 1–18.

Hawari, A., Alamin, M., Alkadour, F., Elmasry, M. & Zayed, T. 2018 Automated defect detection tool for closed circuit television (CCTV) inspected sewer pipelines. *Automation in Construction* **89**, 99–109.

Hüllermeier, E. & Vanderlooy, S. 2010 Combining predictions in pairwise classification: an optimal adaptive voting strategy

and its relation to weighted voting. *Pattern Recognition* **43** (1), 128–142.

Jahanshahi, M. & Marsi, S. 2012 Adaptive vision-based crack detection using 3D scene reconstruction for condition assessment of structures. *Automation in Construction* **22**, 567–576.

Khalifa, I., Aboutabl, A. & Aziz, G. 2014 A new image model for predicting cracks in sewer pipes based on time. *International Journal of Computer Applications* **87** (9), 25–32.

Kohavi, R. 1995 A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International Joint Conference on Artificial Intelligence*, Montreal, Canada.

Lindeberg, T. 2012 Scale invariant feature transform. *Scholarpedia* **7** (5), 10491.

Myrans, J., Everson, R. & Kapelan, Z. 2018 Automated detection of faults in sewers using CCTV image sequences. *Automation in Construction* **95**, 64–71.

Oliva, A. & Torralba, A. 2001 Modelling the shape of a scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision* **42** (3), 145–175.

Panasiuk, O., Hedström, A., Marsalek, J., Ashley, R. M. & Viklander, M. 2015 Contamination of stormwater by wastewater: a review of detection methods. *Journal of Environmental Management* **152**, 241–250.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. & Vanderplas, J. 2011 Scikit-learn: Machine learning in python. *Journal of Machine Learning Research* **12**, 2825–2830.

Sinha, S. & Fieguth, P. 2006 Neuro-fuzzy network for the classification of buried pipe defects. *Automation in Construction* **15** (1), 73–83.

van der Steen, A., Dirksen, J. & Clemens, F. 2014 Visual sewer inspection: detail of coding system versus data quality? *Structure and Infrastructure Engineering* **10** (11), 1385–1393.

WRc 2013 *Manual of Sewer Condition Classification*, 5th revised edn. s.l. WRc Publications, Swindon, UK.

Yan, Z., Chakraborty, D., Misra, A., Jeung, H. & Aberer, K. 2012 *Semantic Activity Classification Using Locomotive Signatures From Mobile Phones*. s.l. (No. EPFL-REPORT-174016).