

Automatic Detection of Sewer Faults Using Continuous CCTV Footage

Joshua Myrans¹, Zoran Kapelan¹, Richard Everson¹

¹University of Exeter, Harrison Building, North Park Road, Exeter, Devon, UK

jm494@exeter.ac.uk

ABSTRACT

The work presented in the paper demonstrates the automatic detection of faults in sewers using CCTV footage. Specifically, it develops the fault detection methodology for application to a continuous video sequence, implementing a new ‘Smoothing’ stage. This new stage implements a Hidden Markov Model and Windowing technique, able to incorporate information from an entire CCTV segment to improve the frame by frame identification. The modified methodology was tested and demonstrated on footage taken from real world surveys in the UK. The results obtained, demonstrate that the extended methodology has improved performance with up to an 8% increase in accuracy and up to 5% increase in area under the ROC curve. Given the ability to produce predictions comparable to those of surveyors, the methodology shows promise for future application in industry.

Keywords: Automatic, Fault Detection, Hidden Markov Model

1 INTRODUCTION

UK water companies manage thousands of kilometres of wastewater networks, all of which require regular inspection in order to prioritise essential maintenance. Currently inspections are performed using CCTV cameras, recording the otherwise inaccessible interior of pipes too small for entry. These cameras are either controlled remotely or pushed through the pipes using a system of rods. Both methods record detailed footage of the pipe’s interior used to identify faults within the system. Footage is often analysed after collection, requiring the entire length of footage to be re-watched and annotated. However, if the camera is remotely controlled, a skilled technician may be able to annotate the footage as it is collected, although this commonly slows the collection as the camera is stationary whilst a fault is annotated. In the UK, annotation is performed using the *WRc Manual of Sewer Condition Classification* [1], which is comparable to the industry standard across the world.

Regardless of collection and annotation method sewer surveys are costly, requiring a team of technicians and expensive hardware. The process is also slow; locating and accessing assets alongside performing the survey can often take days, even for short surveys (<1km of pipe). Companies can also be fined by councils if surveys impact local infrastructure, charging more the longer a survey continues. For these reasons water companies are looking to improve the efficiency of their surveying, reducing costs and enabling the effective management of the rest of their network. This paper presents an automatic fault detection methodology, aimed to reduce the time required to analyse collected footage.

Previous work in the field of automated fault detection have yielded respectable results and utilised many different machine learning techniques. The most notable contributions have been made by Duarn *et al.* [2] and Sinha *et al.* [3], who both applied Artificial Neural Networks (ANN) to identify faults in wastewater pipes. Duran *et al.* [2] augmented the CCTV camera with a laser profiler, mapping the interior of the pipe walls, before identifying the presence of a fault using an ANN. Duran *et al.* [2] continued to predict the category and severity of detected faults, using the same techniques. The work was successfully demonstrated on PVC pipes in lab based experiments. Sinha *et al.* [3] worked on raw CCTV footage, with the aim of detecting structural faults in sewer pipes. The methodology utilised average light intensity to identify areas of interest, before extracting a series of key features, such as: light intensity, shape and size. The measured features are combined with expert knowledge using fuzzy logic and passed to an ANN classifier. Unlike Duran *et al.* [2] this work was demonstrated on real-life CCTV footage collected from flush cleaned pipes. More recently Halfawy & Hengmeechai [4] examined individual frames of CCTV surveys, extracting

Histogram of Ordered Gradient features (HOG) for identified of interest within the frame. Once extracted the HOG features are passed to a Support Vector Machine (SVM), predicting the presence of a fault. This produced good results for the tested categories of fault, although the methodology wasn't demonstrated on a collection of subtler faults, including cracks, deposits and surface damage.

The methodology expanded upon in this paper aims to automatically identify faults using only raw CCTV footage. Continuing on the success of previous work [5][6], this paper aims to extend the 5 stage fault detection methodology to continuous CCTV footage. Previously the methodology had been tested on a library of still images, correctly classifying 92% of the individually analysed frames, however the methodology wasn't quantitatively tested on continuous CCTV footage. This paper aims to demonstrate the methodology on continuous CCTV footage, developing a Markovian structure to incorporate temporal information.

2 FAULT DETECTION METHODOLOGY

As this paper relies heavily on the fault detection methodology [5][6], this will be briefly outlined in order to identify how the improvements will be implemented. Previously the fault detection consisted of 5 stages: 'Extraction', 'Pre-processing', 'Feature Extraction', 'Classification' and 'Interpretation'.

2.1 METHODOLOGY OVERVIEW

The first 'Extraction' stage of the methodology, lifts frames, one at a time, from the raw CCTV footage. Frames should be lifted at a rate relative to the speed of the camera, however the current methodology can analyse every frame of footage in near real time. The second 'Pre-processing' stage aims to reduce the complexity of the classification problem, resizing each frame to a predetermined resolution. In addition, each frame is converted to greyscale, as the extra information provided by colour is often unnecessary in identifying faults. Continuing to reduce the complexity of the classification of the problem, the third 'Feature Extraction' stage extracts a GIST feature descriptor [7] for each frame. This GIST descriptor aims to describe the contents of an image in a much lower dimensional space, whilst maintaining all key features of the original image. The fourth stage, 'Classification', applies a Random Forest (RF) classifier [8], predicting the probability of a frame containing a fault. This machine learning classifier should to be trained on a collection of relevant labelled GIST features prior to application in order to make accurate predictions. Finally given the predictions of the RF classifier, stage 5 'Interpretation' labels the frame based on its predicted probability of containing a fault. This is done using a threshold (normally 0.5), requiring a predicted probability higher than the threshold for a frame to be labelled as containing a fault.

2.2 SMOOTHING

Applying the fault detection methodology to continuous footage adds additional challenge to the classification problem. As the camera is almost always in motion, many frames are noisy, suffering from distortion, making it much harder to identify a frames contents. Another issue is the methodology is no longer being tested on clear examples. Annotating CCTV footage isn't a clear-cut science, even with the *WRc Manual of Sewer Condition Classification* [1] two surveyors can classify a frame of footage differently, as discussed by van der Steen *et al.* [9]. This is most prevalent before and after faults where a faults presence isn't always clear.

The new implementation of the existing fault detection methodology attempts to tackle these issues by wrapping the 5 stage structure in a Markovian framework. To achieve this, stage 5 'Interpretation' is replaced with a new 'Smoothing' stage. The 'Smoothing' stage utilises two techniques, a Hidden Markov Model (HMM) [10] and Windowing (also known as Order Oblivious Filtering) [11].

A HMM uses the predictions of the RF classifier to identify the underlying hidden state of a frame, where the hidden state is either 'normal' or 'faulty'. The probability of the hidden state changing is

governed by a state transition matrix, the parameters of which are calculated using the Baum-Welch algorithm [10]. To incorporate the predicted state of all other frames in the sequence the probability of each frame's state given all frames in the CCTV sequence is finally calculated using the forwards/backwards algorithm [10].

On the other hand, Windowing weights the predicted probability of a frame towards that of its neighbours. Achieved by averaging over a fixed window of neighbouring frames [11], individual frames are classified as 'faulty' or 'normal' according to local probabilities. This improves short term predictions and enables the rejection of isolated inconsistencies in the RF's predictions.

In combination, these two techniques aim to reduce the impact of noise and indecision in continuous CCTV footage. The effect of noise in the recorded footage is reduced by combining predictions across the entire duration of the survey using the global traits of the HMM. Whilst weighting predictions towards local classifications aims to minimise the classifiers indecision over the presence of a fault.

3 CASE STUDY

3.1 Description

To demonstrate the proposed 'Smoothing' stage in the re-worked methodology, both the old and new fault detection methodologies will be applied to three sequences of CCTV footage. The footage used in the case study was provided by UK water company Wessex Water, and was collected during their routine surveys by their technicians and nearby contractors. To train the techniques, a dataset of 1,500 still frames was constructed. These frames were extracted from around 10 hours (5.5km) of CCTV footage, covering pipes of various shapes, a variety of sizes (150 – 1500mm) and various materials, including vitrified clay, concrete and brick. The extracted dataset of frames was labelled according to the *Manual of Sewer Condition Classification* [1], before being shuffled for use as the training dataset. Approximately half of the extracted frames contained faults, with a bias towards faults more common in wastewater networks as shown in Table 1.

Table 1. A table detailing the distribution of faults within the Wessex water dataset

Fault Type	Subtypes	Distribution (%)
Deposits	Attached, Settled	32.7
Joint	Displaced, Open	30.0
Crack	Longitudinal, Circumferential, Multiple, Spiral	14.5
Broken/ Collapsed	-	10.1
Hole	-	6.8
Obstruction	Intruding junctions, Masonry, Protrusions	6.6
Brickwork	Missing mortar, Displaced bricks, Missing Bricks	3.9
Roots	Fine, Tap, Mass	3.7
Infiltration	Running, Gushing	2.8
Surface	-	2.2
Deformation	-	0.4

Alongside the dataset of still images used to train the fault detection methodology, three CCTV sequences have been set aside for testing. These sequences were unused in the creation of the training dataset and contain a variety of common faults. The first two videos are 10 minutes in duration, covering roughly 20m of pipe, whilst the third video is over 30 mins in duration covering 70m of pipe. To fairly compare the both versions of the fault detection methodology, stages 1- 4 are applied to all three CCTV segments, the result of which is independently used by the old 'Interpretation' stage and the new 'Smoothing' stage. When later comparing the techniques, three key performance criteria are considered:

- Accuracy – the number of frames classified in agreement with the surveyor’s annotations. The surveyor’s annotations are regarded as absolute truth during these experiments, however it should be noted that annotations can be inconsistent and prone to human error as shown by van der Steen *et al.* [12].
- Ratio of False positive rate (FPR) to False negative rate (FNR) – the rates at which frames are incorrectly classified by the method. FPR refers to the number of frames classified as faulty, where no fault is present, whilst FNR refers to the number of frames classified as normal where a fault is present. These metrics are key to having confidence in the methodology, a high FPR wastes a technician’s time as frames are unnecessarily flagged. On the other hand, a high FNR is potentially worse, as the methodology ignores faults which would otherwise have been identified by a technician.
- Area under the curve (AUC) – the area under the receiver operating characteristic (ROC) curve, represents the overall performance of the methodology, where the ROC curve visualises the trade-off between FPR and FNR achieved by varying internal classification thresholds. When comparing two methods one clearly outperforms the other, if it has a higher AUC and its ROC curve completely dominates the second method.

Each of the three videos used in the testing of the fault detection methodology have distinctive features, common across routine CCTV surveys. Video 1 contains footage of a particularly unclean vitrified clay pipe, including substantial grease deposits (Figure 1a). This video contains very few structural faults, however displaced joints are present. Video 2 is a much cleaner vitrified clay pipe, containing only minor structural faults, predominantly displaced joints (Figure 1b). Finally, video 3 is much longer and covers the largest variety of faults, including cracks, intruding roots and displaced joints (Figure 1c).

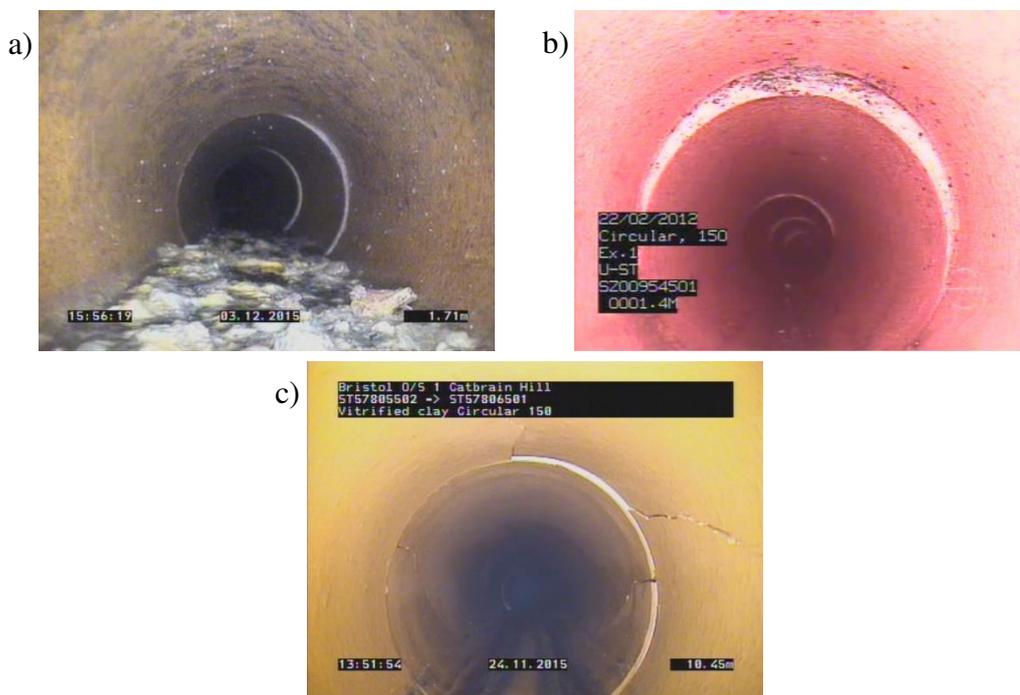


Figure 1. a) an example frame from video 1 b) and example frame from video 2 c) an example frame from video 3

3.2 RESULTS & DISCUSSION

The fault detection methodology was implemented on all three testing video segments, individually applying the old ‘Interpretation’ and new ‘Smoothing’ stage. When applied to video 1, the base

methodology achieved detection accuracy of 74%, whilst the smoothed methodology achieved an improved accuracy of 75%. However, continuing to examine the FPR and FNR rates demonstrated in Table 2, the smoothed methodology achieves a more preferable ratio. The base methodology achieves a low FNR of 10%, but suffers from a very high FPR of 47%, conversely the smoothed methodology has a balance FNR of 27% and FPR of 22%.

Table 2. Confusion matrices for the base and smoothed fault detection methodology on video 1.

Base (0.74)		Truth		Smoothed (0.75)		Truth	
		Fault	Normal			Fault	Normal
Predicted	Fault	0.90	0.47	Predicted	Fault	0.73	0.22
	Normal	0.10	0.53		Normal	0.27	0.78

When applied to the second video sequence, the base methodology achieved detection accuracy of 69%, where the smoothed methodology achieved an accuracy of 75%. Examining the FPR and FNR, the base methodology achieved a FPR of 83% and a FNR of 0%, whilst the smoothed methodology achieved a FPR of 35% and FNR of 18% (Table 3). The extremely low FNR of the base methodology may appear impressive at first but the high FPR means 83% of normal frames are misclassified, again leaving the smoothed methodology with the preferable ratio of FPR to FNR.

Table 3. Confusion matrices for the base and smoothed fault detection methodology on video 2.

Base (0.69)		Truth		Smoothed (0.75)		Truth	
		Fault	Normal			Fault	Normal
Predicted	Fault	1.00	0.83	Predicted	Fault	0.82	0.35
	Normal	0.00	0.17		Normal	0.18	0.65

Applying the methodologies to the final video sequence, the base methodology achieved an accuracy of 64% and the smoothed methodology achieved an accuracy of 72%. The base methodology achieved a FPR of 11% and an FNR of 52%, whilst the smoothed methodology achieved a FPR of 14% and FNR of 37% (Table 4). Again, the smoothed methodology proved its effectiveness, balancing the FPR and FNR rates to a more preferable ratio.

Table 4. Confusion matrices for the base and smoothed fault detection methodology on video 3.

Base (0.64)		Truth		Smoothed (0.72)		Truth	
		Fault	Normal			Fault	Normal
Predicted	Fault	0.48	0.11	Predicted	Fault	0.63	0.14
	Normal	0.52	0.89		Normal	0.37	0.86

Continuing to examine the third performance criteria (AUC), the ROC curve for both versions of the methodology has been plotted for each of the three testing videos, as shown in Figure 2. The ROC curves for the smoothed methodology in all the videos mostly dominate the base

methodology, pushing toward the goal a TPR equal to 1 and FPR equal to 0 with higher AUC values across all smoothed methodologies. The only notable exception is Video 2 (Figure 2b), where the two curves are much closer. This is likely due to displaced joints being the only notable faults within this video footage. This category of fault is well defined, having a definitive start and end, hence the smoothing process provides little benefit. Most impressively ‘Smoothing’ in video 3 showed a 5% increase in AUC, from 77% in the base methodology to 82% in the smoothed methodology

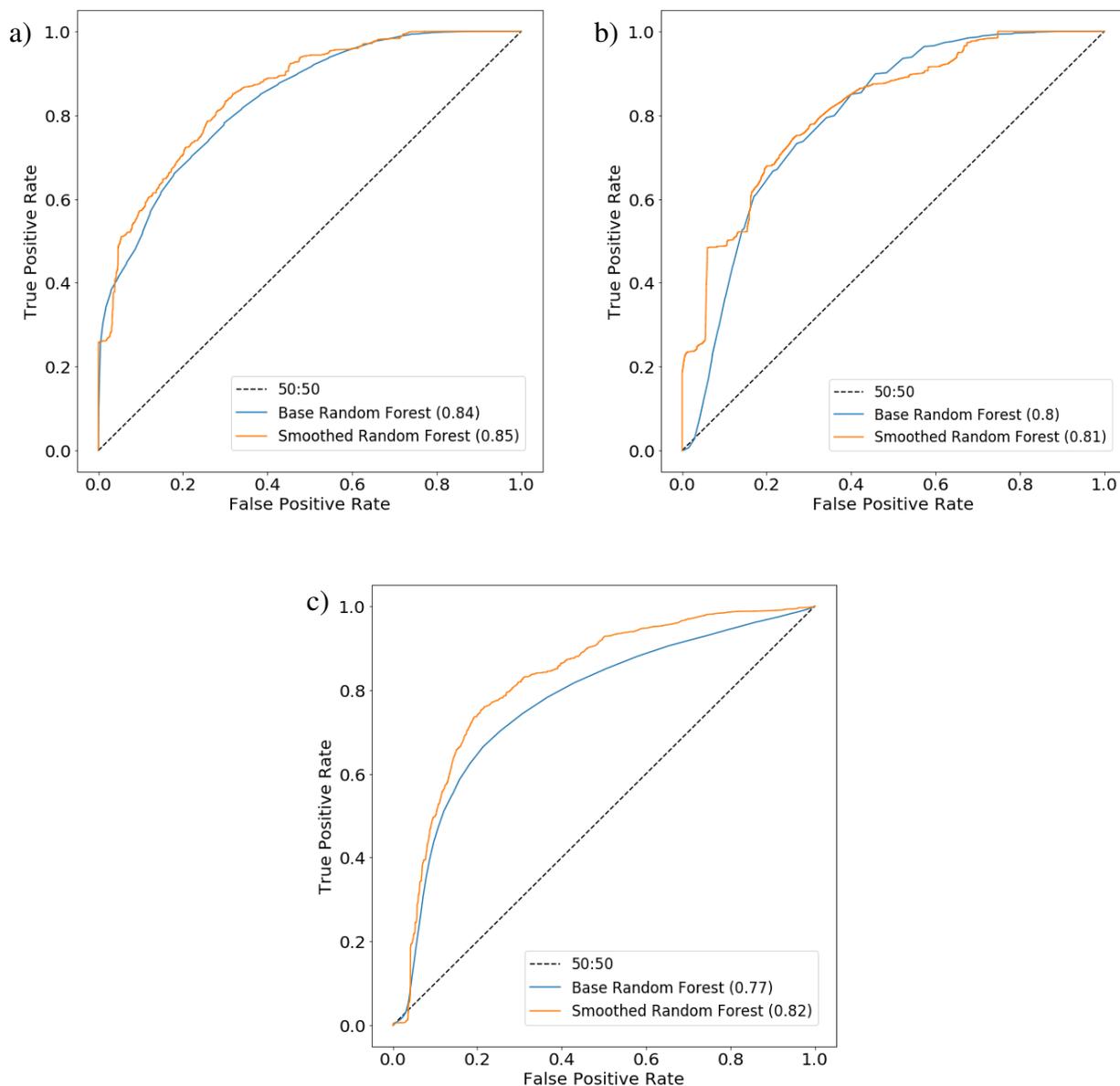


Figure 4. ROC curves comparing the performance of the smoothed fault detection methodology against the base methodology for a) video 1 b) video 2 c) video 3. Where the orange line represents the smoothed methodology, and the blue represents the base methodology. Note the dashed line in represents a 50:50 guess at the contents of each frame.

In addition to examining the quantitative statistics of the methodology, the frame by frame predictions of the methodology were examined, showing the true impact of the ‘Smoothing’ stage. Figure 4 shows a segment of 5,000 frames taken from video 3, and each frame’s assigned state at certain stages of the methodology. When the base ‘thresholding’ methodology is applied, predictions are scattered, influenced by noise and the motion of the camera. After applying the HMM and windowing stages of the smoothed methodology, the improved performance is clear,

decisively categorising blocks of frames as per the surveyor’s annotations. This elimination of isolated misclassifications, clearly defines sequences of footage that are predicted to contain a fault, identifying clear sequences of footage for a technician to analyse.

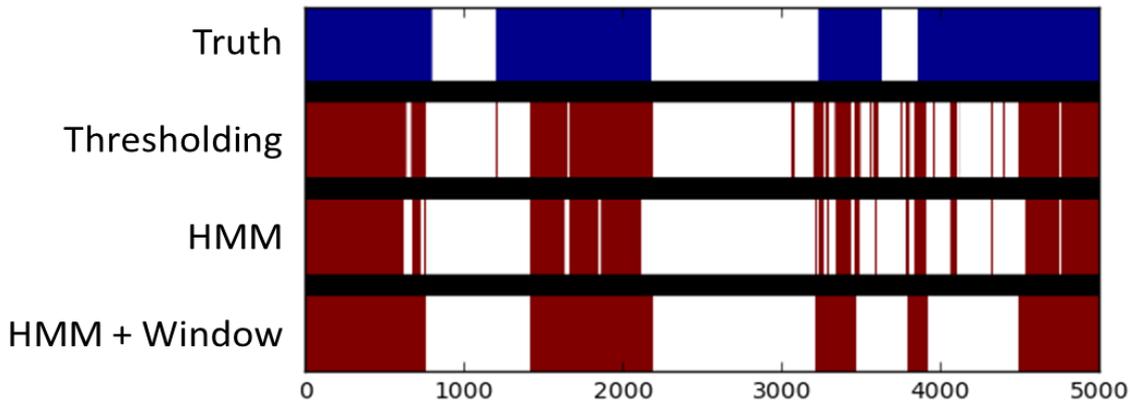


Figure 3. A comparison of the effect of applying different techniques to the Random Forest’s output over a segment of a CCTV survey. Red denotes a ‘faulty’ frame, white denotes a ‘normal’ frame and blue denotes the surveyor’s annotation.

Overall it is clear to see that the addition of the ‘Smoothing’ stage to the fault detection methodology improving on the accuracy and AUC of the base methodology in all tests. ‘Smoothing’ also improves the ratio of false positives to false negatives, achieving a better balance between the two rates. Previously the base methodology had a strong bias, achieving an exceptionally low FPR and very high FNR or vice versa. This is most clearly demonstrated in Table 3, where the base methodology achieved a 0% FPR and 83% FNR. By balancing the two rates a user can be more confident in the predictions of the methodology, as there is a higher chance of both predicted classes being accurate. It should still be noted that the smoothed predictions still do not quite match the surveyor’s annotations, most obviously in Figure 3. The model is not perfect, certainly accounting for a share of the discrepancy, but as previously mentioned, the art of annotating surveys can be very subjective. Different technicians identify faults at varying times, often merging similar neighbouring faults together [9]. However, the results demonstrated across all experiments in this paper proved an ability to flag at least sections of faults within the CCTV footage. As this methodology is designed to speed up the process of fault detection, working alongside a technician, footage neighbouring a flagged fault could be explored to ensure an accurate analysis.

4 CONCLUSIONS

The fault detection methodology, alongside the new ‘Smoothing’ stage show promise for application in industry, given the ability to quickly and accurately achieve results comparable to technicians. Proving its usefulness in balancing FPR and FNR, whilst improving accuracy, smoothing will likely become a permanent addition to the fault detection methodology. The implementation of the ‘Smoothing’ stage also groups faulty frames together, clearly defining segments of footage for further analysis, speeding up the surveying process. With further testing on a larger scale, the entire methodology’s effectiveness can be ensured, able to provide assistance in the detection of sewer faults.

Future work in this area aims to extend the methodology, categorising detected faults, according to the specifications in the *WRc Manual of Sewer Condition Classification* [1]. In addition, the extra information provided by the structure of the Random Forest classifier can be interpreted to identify areas of interest within a frame. These areas of interest, key to classifying the frame could lead to the location of a fault, which could be useful to the technician annotating the frame. Ultimately,

these technologies will be combined to produce a universal decision support tool, with the goal of dramatically reducing the time required to analyse sewer surveys.

5 REFERENCES

- [1] WRc plc, 2013. *Manual of Sewer Condition Classification. 5th Revised edition*. WRc Publications.
- [2] Duran, O., Althoefer, K. and Seneviratne, L.D., 2007. *Automated pipe defect detection and categorization using camera/laser-based profiler and artificial neural network*. Automation Science and Engineering, IEEE Transactions on, 4(1), pp.118-126.
- [3] Sinha, S.K. and Fieguth, P.W., 2006. *Neuro-fuzzy network for the classification of buried pipe defects*. Automation in Construction, 15(1), pp.73-83.
- [4] Halfawy, M.R. and Hengmeechai, J., 2014. *Automated defect detection in sewer closed circuit television images using histograms of oriented gradients and support vector machine*. Automation in Construction, 38, pp.1-13.
- [5] Myrans, J., Kapelan, Z. and Everson, R., 2016a. *Automated Detection of Faults in Wastewater Pipes from CCTV Footage by Using Random Forests*. Procedia Engineering, 154, pp.36-41.
- [6] Myrans, J., Kapelan, Z., Everson, R. and Britton, J., 2016b. *Using Support Vector Machines to identify faults in sewer pipes from CCTV surveys*. CCWI 2016, Electronic proceedings
- [7] Oliva, A. and Torralba, A., 2001. *Modelling the shape of the scene: A holistic representation of the spatial envelope*. International Journal of Computer Vision, 42(3), pp.145-175.
- [8] Breiman, L., 2001. *Random Forests*. Machine learning, 45(1), pp.5-32.
- [9] van der Steen, A.J., Dirksen, J. and Clemens, F.H., 2014. *Visual sewer inspection: detail of coding system versus data quality?*. Structure and Infrastructure Engineering, 10(11), pp.1385-1393.
- [10] Rabiner, L.R., 1989. *A tutorial on hidden Markov models and selected applications in speech recognition*. Proceedings of the IEEE, 77(2), pp.257-286.
- [11] Yan, Z., Chakraborty, D., Misra, A., Jeung, H. and Aberer, K., 2012. *Semantic activity classification using locomotive signatures from mobile phones* (No. EPFL-REPORT-174016).