Automation in Construction 95 (2018) 64–71

Contents lists available at ScienceDirect

Automation in Construction

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

Automated detection of faults in sewers using CCTV image sequences[☆]

Joshua Myrans[a,*], Richard Everson[b], Zoran Kapelan[c]

[a] University of Exeter, Harrison Building, North Park Road, Exeter, Devon, UK.
[b] Department of Computer Science, Harrison Building, North Park Road, Exeter, Devon, UK.
[c] Centre for Water Systems, University of Exeter, Harrison Building, North Park Road, Exeter, Devon, UK.

ARTICLE INFO

Keywords:
Fault detection
GIST
Hidden Markov model
Random forest
Sewer pipe
Support vector machine

ABSTRACT

Routine CCTV surveys are vital to the effective maintenance of wastewater networks, but their time-consuming nature makes them very expensive. We present a methodology capable of automatically detecting faults within recorded CCTV footage, aiming to improve surveying efficiency. The procedure calculates a feature descriptor for each video frame, before using a machine learning classifier to predict the contents of individual frames. The sequence of predictions is then smoothed using a Hidden Markov Model and order oblivious filtering, incorporating information from the entire sequence of frames. This technique has been demonstrated on footage collected by the Wessex Water, achieving a detection accuracy of over 80% on still images. Furthermore, temporal smoothing on continuous CCTV footage improved false negative rate by more than 20%, to achieve an accuracy of 80%. This last step enables the method to compete with the performance of trained technicians, showing promise for application in industry.

1. Introduction

1.1. Motivation and aim

UK water companies are each responsible for tens of thousands of kilometres of wastewater networks. In order to provide the best service to their consumers and comply with government regulations these networks require regular maintenance. To efficiently allocate the limited resources available for refurbishment, routine surveys are required.

Identifying and annotating faults in sewer pipes is extremely important for a number of reasons. Different faults in sewers can act as seed points for initiating blockages (e.g. root intrusions, deposits, pipe displacements) or in more severe cases, result in sewer collapses (e.g. via different type cracks in sewer walls). In less severe cases, the existence of cracks and similar faults is likely to result in increased infiltration into sewers thus reducing the conveying capacity of a system during storm events. All this, in turn, is likely to result in flooding and/or pollution incidents thus potentially exposing untreated wastewater to customers and the environment (e.g. aquatic life in rivers). In the UK alone, in 2014–15 there were, on average, two serious and 71 less-severe sewerage related pollutions incidents for every 10,000 km of sewers [1]. This is not to say all the above incidents were caused by sewer defects: pollution events can also be attributed to under-designed and misused networks. However, it is clear that the number of pollution events will reduce given better maintained networks.

Currently surveying a wastewater network is a time consuming and expensive process, requiring qualified human interaction at all stages. Surveys often involve propelling a CCTV camera through the network recording the pipe's interior for later analysis, which is achieved in one of two ways. The first, and most common approach, is for a team of engineers to force a small camera through the pipe on the end of a semi-rigid cable. Alternatively, a remote-controlled pipe inspection gadget (PIG) can be driven through the network, collecting footage of the pipe's interior with its on-board camera. Although not as common as a cable fed camera, PIGs often yield superior surveys, because they allow the operator to remotely control the camera and further investigate faults. Additional technologies can also be equipped to a PIG, although they are often expensive and not currently common in practice. Technologies include: laser profilers, which can accurately measure the internal dimensions of a pipe [2] and 3D cameras, capable of record 360-degree panoramic footage of the pipes interior [3].

Surveys performed using any of these techniques are often analysed

☆ This work was supported by Engineering and Physical Sciences Research Council in the UK via grant EP/L0116214/1 awarded for the Water Informatics, Science and Engineering (WISE) centre for doctoral training.

The data that support the findings of this study are available on request from the corresponding author, JM. The data are not publicly available as it is property of Wessex Water Services Ltd.

* Corresponding author.
E-mail address: jm494@exeter.ac.uk (J. Myrans).

https://doi.org/10.1016/j.autcon.2018.08.005
Received 6 March 2018; Received in revised form 13 July 2018; Accepted 10 August 2018
0926-5805/

after collection (i.e., offline). This requires an engineer to re-watch the footage, annotating any faults with key details as per the *Manual of Sewer Condition Classification* [4]. However, if a surveyor has a live feed to the recording device (normally via a PIG) the footage can be analysed as it is recorded (i.e., online). As the surveyor will be controlling the PIG and annotating the footage simultaneously, this process is often faster than offline analysis. However, online annotation slows the recording process, leaving the PIG stationary whilst the operator makes notes. This longer recording time can be costly, especially if it causes disruption to local infrastructure.

This work describes a new automated method that makes use of advanced machine learning and image processing techniques to automatically identify faults of all types in CCTV sewer surveys. The proposed methodology was designed to work using only industry standard CCTV footage—rather than specialised hardware—with the goal of producing a decision support technology for use in the 'real world', able to assist in surveys performed across the water industry. Doing so will (a) greatly reduce the time and costs associated with current sewer surveys by requiring expert human input only for analysis of flagged faults and (b) improve the consistency and reliability of identified and reported faults.

1.2. Related work

The field of automatic sewer fault detection has a rich background, consistently applying modern machine learning techniques to achieve promising results. Early work in the field by Duran et al. [3,5,6] and Sinha et al. [7] applied Artificial Neural Networks (ANNs) to the problem. Duran et al. retrofitted a CCTV camera with a laser profiler, passing precise internal measurements to an ANN to identify structural faults. On the other hand, Sinha et al. identified key features within the CCTV footage, applying Fuzzy Logic to characteristics such as light intensity, shape and size. These fuzzy features were again passed to a trained ANN to identify cracks within CCTV footage.

Crack detection technology has continued to flourish since Sinha et al.'s work: Khalifa et al. [8] used a Canny edge detector in a Markovian framework to identify cracks over multiple CCTV frames. Similarly, Halfawy and Hengmeechai [14] detected edges in frames, using the Sobel method, before filtering the image with the Hough transform and a suite of morphological operations to identify cracks in pipes. Opting not to use edge detectors, Jahanshani and Masri [9] developed a 3D scene structure to help identify cracks and their depths in sewer pipes. These structures were classified using an ANN and Support Vector Machine (SVM) to determine the presence of cracks. Most recently Chen et al. [10] used local binary patterns (LBP) with a SVM and Bayesian decision theory to detect cracks in metallic surfaces. Even though the technique was not directly applied to sewer inspections, the methodology may be transferable.

Beyond pure crack detection, Guo et al. [11] developed a technique based on frame differencing: by comparing adjacent frames in video footage, Guo et al. identified areas of interest by looking for abrupt changes in the frames' contents. Guo et al. [12] continued their work, using Scale Invariant Feature Transform (SIFT) features to identify patterns in a video sequence, choosing to use SIFT features for their robustness to changes in scale, illumination and rotation. More recently Halfawy and Hengmeechai [13] tracked the motion of the CCTV camera using optical flow and a SVM. Following the motion of the camera (controlled by the operator), enabled regions of interest to be inferred as the operator looked for faults. Finally, Halfawy and Hengmeechai [15] used HOG (Histogram of Gradients) features together with a SVM to identify local regions of interest within frames of CCTV footage. HOG features were chosen as they are designed to capture the shape of local regions and proved effective, demonstrating an accuracy of 86% when detecting tree root intrusions.

1.3. Contribution

The methodology presented in this paper was developed from the ground up taking inspiration from previous work, to develop a holistic fault detector. The detector is holistic in the sense that, unlike other similar approaches, it uses a single technique capable of detecting most types of faults encountered in sewer pipes. As described above, previous work has been concentrated on detecting individual faults, most notably cracks. However, there are many distinct types of fault defined by the *Manual of Sewer Condition Classification* [4] which, for effective everyday use, would require the development of a library of these specialised detectors. Those methodologies which apply comparable techniques, reported accuracies of at least 80%, however they weren't demonstrated on continuous CCTV footage and utilised ANNs which require a complex calibration process [16].

This paper presents an objective assessment of two classification methods, Random Forests and Support Vector Machines [17,18] applied to classify GIST feature descriptors to detect general faults in sewer pipes from still images. In addition, the method is demonstrated on continuous CCTV footage taken from industry surveys. Using a Hidden Markov Model (HMM) and an order oblivious filter allows additional information from neighbouring frames to be used to obtain more accurate predictions, dramatically improving the False Negative Rate (FNR). In essence this minimises interference from the camera's movement during the classification of continuous footage, a problem which is less prevalent when classifying individual frames. The overall method is computationally fast enough to permit automatic, online, near real-time detection of general faults in sewer pipes.

2. Fault detection methodology

2.1. Overview

The fault detection methodology applies modern image processing and machine learning techniques to raw CCTV footage, taken directly from industry surveys. The method is data-driven in the sense that relevant image characteristics and classifier parameters are learned from a database of images that have been labelled as containing a fault or normal. It is therefore assumed that such a database is available. This database can constantly grow, as new faults are identified and verified by a technician. These extra samples should lead to an improved detection accuracy, when the methodology is retrained on its corrected misclassifications. The training process, although relatively time consuming (a few hours) for a computer does not required human intervention and so can be performed periodically overnight.

As illustrated in Fig. 1, the online method examines each frame in a survey in turn, predicting the presence of a fault in each frame in turn, before smoothing predictions across neighbouring frames. This structure can be intuitively broken down into four stages: 'Frame Extraction & Pre-processing', 'Feature Extraction', 'Classification' and 'Smoothing', each of which is described below. Operational values for parameters are given in Table 1.

2.2. Frame extraction & pre-processing

In order to process frames individually, each frame is extracted from the raw CCTV footage as an RGB image. Frames are generally recorded at a nominal rate of 25 frames per second (fps) and since, as shown in Section 3.2.2, the computational effort to process each frame is light, all available frames are extracted. Whilst it would be possible to omit some frames, excessive reduction of the sampling frequency will detrimentally impact performance. In the work presented here, all recorded frames were used, as shown in Table 1.

Each extracted frame is converted to greyscale and resized to a lower resolution R. Experimentation showed that retaining colour information degrades classification performance because of the wide