Automation in Construction 91 (2018) 273–283

Contents lists available at ScienceDirect

Automation in Construction

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

Automated defect classification in sewer closed circuit television inspections using deep convolutional neural networks

Srinath S. Kumar[a,*], Dulcy M. Abraham[a], Mohammad R. Jahanshahi[a], Tom Iseley[b], Justin Starr[c]

[a] Lyles School of Civil Engineering, Purdue University, West Lafayette, Indiana, United States
[b] Department of Civil Engineering, Louisiana Tech. University, Ruston, LA, United States
[c] RedZone® Robotics, Pittsburgh, PA, United States

ARTICLE INFO

Keywords:
Wastewater pipelines
Underground infrastructure
Automation
Robotics
Closed circuit television (CCTV)
Convolutional neural networks
Condition assessment
Deep learning
Artificial intelligence

ABSTRACT

Automated interpretation of sewer CCTV inspection videos could improve the speed, accuracy, and consistency of sewer defect reporting. Previous research has attempted to use computer vision, namely feature extraction methods for automated classification of defects in sewer CCTV images. However, feature extraction methods use pre-engineered features for classifying images, leading to poor generalization capabilities. Due to large variations in sewer images arising from differing pipe diameters, in-situ conditions (e.g., fog and grease), etc., previous automated methods suffer from poor classification performance when applied to sewer CCTV videos. This paper presents a framework that uses deep convoluted neural networks (CNNs) to classify multiple defects in sewer CCTV images. A prototype system was developed to classify root intrusions, deposits, and cracks. The CNNs were trained and tested using 12,000 images collected from over 200 pipelines. The average testing accuracy, precision and recall were 86.2%, 87.7% and 90.6%, respectively, demonstrating the viability of this approach in the automated interpretation of sewer CCTV videos.

1. Introduction

The United States has over 800,000 miles of public sewage pipes and 500,000 miles of private sewer laterals [1]. While municipalities have invested in expanding sewer systems to meet growth and treatment plant upgrades, a relatively smaller investment has gone into sewer rehabilitation [2,3]. The lack of funding towards wastewater infrastructure rehabilitation is highlighted in the Clean Watersheds Needs Survey, which estimated the wastewater and stormwater treatment and collection requirements for the US at $271 billion, as of January 1, 2012 [4]. As a result, municipalities across the US face the problem of aging sewer infrastructure in dire need of repair, rehabilitation or renewal. Inadequate maintenance causes inflow and infiltration, sanitary sewer overflows, and sinkholes, which not only threaten human health, but also tend to be expensive to correct. According to the United States Environmental Protection Agency (EPA), inflow and infiltration, which are often caused by cracks in sewer walls, root intrusions, and leaky manholes, cost municipalities an additional $2 to $5 per thousand gallons of sewage [5]. The EPA also estimates between 23,000 and 75,000 sanitary sewer overflows each year in the US, which release large quantities of untreated sewage into the environment, exposing humans to a variety of illnesses [6]. Sinkholes caused by sewer collapses cause loss of life and extensive damage to property. For instance, the cost to repair the Fraser sinkhole in Michigan in December 2016 is estimated at over US $78 million. Municipalities have found it economically viable to apply reactive strategies, repairing when failures occur, however, this approach is estimated to become less viable as sewer systems age and the infrastructure funding gap continues to increase [7].

The planning of maintenance, renewal and rehabilitation activities requires accurate information about the condition of sewer pipes. Over the past 40 years, municipalities in North America have been using CCTV as the primary technique for inspecting the internal surface of non-entry sewers [8]. Most CCTV systems employ a pan and tilt camera to record a video of the pipe's interior surface. The drawback of CCTV systems is that the operator has to stop and turn the camera every time a region of interest (ROI) is encountered. As a result, inspection rates are low, causing municipalities to inspect only a fraction of their network due to budget restrictions [9]. Another drawback of CCTV sewer inspections is inconsistency in defect reporting [10]. While these inconsistencies can be alleviated through training and the use of standardized reporting formats, such as the Pipeline Assessment Certification Program (PACP), the operator's experience, skill and biases are known to significantly affect the inspection report. In the last decade,

* Corresponding author.
E-mail address: kumar256@purdue.edu (S.S. Kumar).

https://doi.org/10.1016/j.autcon.2018.03.028
Received 31 October 2017; Received in revised form 27 January 2018; Accepted 20 March 2018

there has been an emergence of multi-sensor inspection technologies, combining wide angled lenses, side-scanners, laser profilers, sonar, and electrical methods [11–13]. Although these technologies have demonstrated significant improvements in sewer condition assessment, at present CCTV remains by far the most widely used method in North America. The popularity of CCTV inspection may be in part due to the lower upfront cost, relative ease of use, and municipalities' and contractors' familiarity with this method of inspection. Developing methods to improve the speed, consistency and reliability of existing CCTV methods would thus be of enormous benefit to sewer condition assessment.

An automated system for extracting sewer condition information from CCTV videos would be beneficial to both short-term (sewer inspections) as well as long-term (deterioration modelling) asset management strategies. The benefits to sewer inspection include: (1) minimization of inspection errors due to fatigue, biases and differing skill levels of operators, (2) detection of defects which go unnoticed to the human eye, and (3) enabling quick off-site reviews and quality control of CCTV videos to ensure accuracy of the inspection process [14]. Automated defect classification algorithms can also be incorporated into existing CCTV inspection reporting tools to facilitate easy adoption by municipalities and contractors.

Automated systems for extracting sewer condition information can also be used to re-evaluate large volumes of archived inspection videos, which would be uneconomical to pursue by human defect coders. The re-evaluation of archived videos would facilitate the creation of large datasets of sewer condition information (e.g., structural condition, types and locations of defects, etc.), which is critical to the development of reliable deterioration models [2].

2. Related studies

The algorithm proposed in this paper builds on previous work in automated defect classification in sewer images, as well as on recent advances in deep learning for pattern recognition. This section reviews related research studies.

2.1. Automated feature and defect classification in sewer CCTV inspections

Many research attempts have been made to develop automated systems for interpreting CCTV sewer inspection videos by combining computer vision and image processing techniques [2,8,14–22]. In most of these studies, images of the sewer pipe are classified as either containing or not containing (ROIs). ROIs refer to defects and features such as deposits, roots, connections, joints, etc., which need to be identified in images to facilitate condition assessment. The classification of images in previous studies is based on a feature extraction approach where manually-specified pixel signatures such as edges, gradients, textures, etc., are used to classify images.

Xu et al. [21] developed a framework for the structural assessment of sewers through the classification of joint deformations, distortions, and deposits. Their framework used image processing techniques such as edge detection, binary image thresholding, thinning, and the discrete Fourier transform to extract the joint structure and compute its deformation [21]. Moselhi and Shehab developed an early framework to classify surface defects in sewer images [22]. Their framework used image processing techniques such as smoothing and edge detection to extract image features, and shallow neural networks to classify the images [22]. Yang and Su [17] used the discrete wavelet transform and co-occurrence matrices as a basis to classify images. Their method, which used a support vector machine (SVM) classifier, yielded an average accuracy of 60% in identifying defects such as open joint, deposit, broken pipe and fracture [17].

Other research studies have proposed a two-step methodology where ROIs are first segmented (or isolated) from the background, and later classified using feature extraction methods [8,14–17,19,20,23]. Segmentation of ROIs is performed to reduce the search space and computational complexity of feature extraction. Guo et al. [20] proposed a method to automatically identify ROIs in pipe inspection videos, through a change detection-based approach. Their method used the concept of frame differencing, where each inspection image is subtracted from a reference image of a healthy section [20]. Based on the changed and unchanged pixels, the images were then segmented into ROIs. The ROIs were input to feature extraction and classification modules, which were used to identify the presence of defects. However, the method proposed by Guo et al. [20] worked best with images of clean surface walls by using fixed angle cameras, and when this condition was not met, a large number of false alarms could be introduced, significantly increasing the processing time. Halfawy and Hengmeechai [14,23] proposed a morphological method, which used differences in brightness of objects in focus as a basis to segment ROIs. Their method used histograms of oriented gradients (HOG) and an SVM classifier, trained with 1000 images to classify the ROIs as containing or not containing defects [23]. Su and Yang [19] proposed an edge detection based morphological approach for segmenting images of deposits and open joints. Their study showed that morphological segmentation based on edge detection (MSED) outperforms morphological opening top-hat operation (OTHO) and closing bottom-hat operation (CBTHO) in crack detection, whereas OTHO outperforms CBHO and MSED in the detection of open joints [19].

The use of morphologies and feature extraction approaches limits the generalization capability of previously developed automated systems. Generalization capability of a classifier can be defined as the ability to classify images that exhibit significant variations (i.e., in terms of shape, color, texture, etc.) from the images used for training. For instance, morphological operations require structuring elements (e.g., simple shapes used to interact with images) to be manually calibrated through repeated trials on training images. As a result, structuring elements which may work well with training images captured under particular conditions (e.g., focal length of camera, distance between object and camera, focal length of lens, illumination conditions, etc.) may not be optimal for images captured under different conditions. Furthermore, morphological approaches are susceptible to generating false positives in sewer pipes that have variations in internal surface colors due to surface staining, relining of pipes, change of material, etc. As a result, morphological approaches cannot be successfully used to extract ROIs in images that differ significantly from the training images, which the structuring elements are based on.

Feature extraction methods, which use pre-engineered (or manually-specified) features, have worked well in areas of pattern recognition, such as face detection, pedestrian tracking, etc. However, classifying images based on pre-engineered features results in a poorer generalization capability than recent deep learning-based automatic feature extractors [24]. As a result, the use of feature extraction methods for classifying images leads to a reduction in performance if the images used for testing vary significantly from the images used for training. Traditional feature extraction methods thus lack the generalization capability to deal with sewer CCTV images that are known to exhibit large variations arising from differences in pipe geometry, materials, nature of defects, presence of internal linings, camera specifications, etc. Previous automated sewer CCTV image classification methods have yielded high defect classification accuracies when applied to small datasets of images collected from a few sources. For instance, the method proposed by Halfawy and Hengmeechai [23] yielded an average classification accuracy of 86% in classifying root intrusion defects, when tested on a set of 100 sewer CCTV images collected from Regina and Calgary, in Canada. However, in order to develop an automated system for interpreting archived inspection videos, the classification performance, as measured by the accuracy, precision and recall of the classifier, should be tested on large datasets of images collected from multiple pipeline inspections.