Automation in Construction 38 (2014) 1–13

Contents lists available at ScienceDirect

Automation in Construction

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

Automated defect detection in sewer closed circuit television images using histograms of oriented gradients and support vector machine

Mahmoud R. Halfawy *, Jantira Hengmeechai [1]

Infrastructure Data Solutions, Inc. (IDS), 2 Research Drive, Suite 150-E, Regina, SK S4S 7H9, Canada

ARTICLE INFO

Article history:
Accepted 19 October 2013
Available online 30 November 2013

Keywords:
Automated defect detection
Closed circuit television inspection
Pipe defects
Computer vision
Pattern recognition
Histograms of oriented gradients
Machine learning
Support vector machine

ABSTRACT

Condition assessment of municipal sewer pipes using closed circuit television (CCTV) inspections is known to be time consuming, costly, and prone to errors primarily due to operator fatigue or novicity. Automated detection of defects can provide a valuable tool for ensuring the quality, accuracy, and consistency of condition data, while reducing the time and cost of the inspection process. This paper presents an efficient pattern recognition algorithm to support automated detection and classification of pipe defects in images obtained from conventional CCTV inspection videos. The algorithm employs the histograms of oriented gradients (HOG) and support vector machine (SVM) to identify pipe defects. The algorithm involves two main steps: (1) image segmentation to extract suspicious regions of interest (ROI) that represent candidate defect areas; and (2) classification of the ROI using SVM classifier that was trained using sets of HOG features extracted from positive and negative examples of the defect. Proposed algorithm is applied to the problem of detecting tree root intrusions. The performance of linear and radial basis function SVM classifiers evaluated. The algorithm was tested on a set of actual CCTV videos obtained from the cities of Regina and Calgary in Canada. Experimental results demonstrated the viability and robustness of the algorithm.

1. Introduction

Municipalities around the world are facing unprecedented challenges in managing aging sewer networks in the face of increasing demands, declining budgets, and increasing environmental and health requirements. Accurate information on the condition of the pipes is the key for effective management and renewal planning of the pipe inventory. Condition information on sewer pipes is commonly collected using closed circuit television (CCTV) inspection systems [40]. CCTV systems involve the use of a remote-controlled forward-looking pan, tilt, and zoom camera mounted on a robot crawler that travels between two manholes along the pipe axis. The CCTV equipment is operated by certified operators who are trained to effectively control the camera to examine various regions inside the pipe, interpret videos, and deduce and record an accurate assessment of the type and severity of the observed defects.

CCTV inspections have a number of limitations. CCTV images only show sewer internal surface above the waterline and do not provide any information on the wall structural integrity or information on the surrounding soil [11]. The quality of CCTV videos, and hence the accuracy of the assessment, relies to a large extent on the camera specifications, camera orientation and vantage point set by the operator, and possible background noise due to non-uniform illumination conditions inside the sewer [38]. To overcome inherent limitations of CCTV systems, a number of sewer inspection technologies have emerged over the last two decades. Examples of these technologies include digital side scanners, sonar and laser scanning, acoustic monitoring systems, and ground penetrating radar. Feeney et al. [11] provided a comprehensive review of the most commonly used sewer inspection technologies, and discussed the advantages and disadvantages of each technology. However, these technologies still have relatively limited use in municipalities, and CCTV systems remain, by far, the most widely used sewer inspection technology. CCTV is also expected to remain an important inspection tool in any condition assessment program for wastewater collection systems [48].

CCTV inspections are known to be costly, time consuming, labor intensive, and error prone [13]. The quality and accuracy of the inspection and assessment process largely depend on the operator's experience and skill level [53]. The tedious and subjective nature of the process presents many challenges in ensuring the consistency and accuracy of the obtained assessment. Also, operators may be subject to fatigue due to lengthy inspection sessions, which could lead to erroneous assessment of the sewer condition. Therefore, CCTV inspections are often required to be reviewed and audited off-site [6], which would add to the cost and time of the condition assessment process.

An automated defect detection system would be an extremely valuable tool for ensuring the quality and accuracy of the CCTV inspections

* Corresponding author at: Research Drive, Suite 150-E, Regina, SK S4S 7H9, Canada. Tel.: +1 306 790 1415.
E-mail addresses: Mahmoud.Halfawy@infrastructure-data-solutions.com (M.R. Halfawy), Jantira.Hengmeechai@infrastructure-data-solutions.com (J. Hengmeechai).
[1] Tel.: +1 306 790 1415.

0926-5805/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.autcon.2013.10.012

and condition assessment information. The system could be used to support off-site reviews and quality control processes. Also, the system could be used to re-evaluate archived CCTV videos to extract historical condition data often required for developing models that predict the long-term deterioration of the pipes [15,39,47]. Furthermore, an automated defect detection system could be integrated into existing CCTV software tools to support the inspection process by providing feedback and alerting the inspectors with the possible location of defects, thus avoiding possible misinterpretation or missing of defects due to operator fatigue or novicity. The system may also become one of the components of multi-sensor autonomous inspection robots [11,24], where data from the CCTV camera are integrated with data from other sensors (e.g., sonar and laser scanners) to provide a more complete interpretation of the pipe condition.

The development of automated defect detection techniques to support sewer inspections has been an active area of research during the past decade. However, Guo et al. [13] noted that a comprehensive method for automated defect detection in sewer inspection images has yet to be established. Most of the proposed techniques have, in general, employed computer vision and pattern recognition algorithms to identify and locate important features and defects in the images. The vast majority of these techniques were designed to work with 2D unfolded sewer surface images obtained using side scanning systems such as the sewer scanning evaluation technology (SSET), the iPEK DigiSewer system, or the IBAK PANORAMO system [22]. However, given the limited use of side scanning systems in municipalities relative to the use of CCTV systems, there is an increasing need to develop automated defect detection tools for conventional CCTV systems, which currently represent the most widely used tool for sewer inspection.

2. Related work

The algorithm proposed in this paper builds on the work in the field of pattern recognition, as well as on previous work in automated defect detection in sewer pipe inspection images. There is a considerable body of work in these two fields in the literature. This section reviews some of the most relevant research to the current study.

2.1. Pattern recognition and object detection

Pattern recognition methods generally focus on defining distinctive image features that would enable correct detection of the object of interest with high probability [9]. Simple object detection approaches that rely solely on pixel intensity values may not be robust enough in most practical applications, due to the noise often encountered in real environments (e.g., changes in illumination conditions, camera viewpoint, etc.). Therefore, more recent approaches are often based on the identification of a set of unique and distinctive features of object's texture and geometry [9,32].

In these approaches, images are mapped from their pixel space into a set of representative features that can uniquely describe the key visual characteristics of the object of interest. Viola and Jones [51] indicated that using image features for object detection have the advantage of encoding domain knowledge using limited sets of training data on the object of interest. Also, computation of image features would generally be more efficient and faster than the expensive methods that rely on pixel intensities for object matching. In essence, the features would represent a signature or a model of the object of interest that can be used to compare the object's image with different regions in the image until a close match is found.

Researchers have developed a wide range of features that can describe the visual characteristics of objects. An overview of the commonly used methods in pattern recognition can be found in [9]. In general, image features often include color, texture, and shape characteristics of the objects. Image textural features [32] have been used extensively in object detection applications. In this approach, textural features are estimated at each pixels and a set of statistical measures are computed from the distribution of the local features in the image. The most commonly used texture features are the ones computed from the gray-level co-occurrence matrix (GLCM). Haralick et al. [16] proposed a set of 28 such features, including inverse difference moment, energy (or angular second moment), contrast, correlation, variance, and entropy. These features are subsequently used to identify specific patterns in an image to support the segmentation and classification processes.

Haar wavelet features, also called Haar-like features, have been used in many pattern recognition applications to increase the efficiency of the computation and the accuracy of the classification compared to the use of raw pixel values [25]. These features are used to classify regions in an image by calculating differences in image intensities among adjacent rectangular regions inside a detection window. Papageorgiou and Poggio [37] used Haar wavelet features and a polynomial support vector machine (SVM) classifier for the detection of faces, people, and cars in static images. Viola and Jones [51] proposed a technique for fast computation of the features, using a representation called "integral image," by calculating the difference of the sum of pixel intensities inside rectangular areas that are scanned over different positions on the original image. They used the Haar wavelet features and a learning algorithm based on a cascade of AdaBoost classifiers to solve the face detection problem. AdaBoost (or adaptive boosting) is a learning technique that constructs a "strong" classifier (i.e., with high classification accuracy) by successively building and combining weaker classifiers. Viola et al. [52] later extended their approach by combining the motion information with the appearance information to detect moving pedestrians in videos.

One of the most commonly used object detection approaches was developed by Lowe [26,27]. This approach is based on the use of scale invariant feature transform (SIFT) features, where a set of scale-invariant keypoints are detected and arrays of orientation histograms at these keypoints are used as descriptors. Using a reference set of images, the SIFT features are first computed and stored in a database. These stored features are then compared with those of a new image to find possible matching features based on the Euclidean distance of the feature vectors. Several researchers have further generalized the SIFT concept to edges. For example, Belongie et al. [1] defined signatures of 2D objects based on edges, called shape contexts, to estimate shape similarity and correspondences. Also, Mikolajczyk et al. [31] proposed an approach based on local edge features that are invariant to affine transformations. Their approach was successfully used to recognize 2D objects with holes and tubular components in noisy and cluttered backgrounds.

More recently, Dalal and Triggs [7] proposed an efficient object detection approach that, in essence, builds on the SIFT and edge-based approaches, and employs the concept of orientation-based histogram features [12]. Their approach involves using contrast-normalized histograms of oriented gradients, or HOG, features computed over a dense grid to characterize the local appearance and shape of the object of interest. A linear SVM classifier was then used to classify objects based on the HOG features. The technique was successfully applied to the problem of human detection in static images. The present study applies the Dalal and Triggs' approach on defect detection in CCTV images of sewer pipes.

In general, the set of image features used in a particular object detection application are often required to meet four main criteria [9,27,32]: (1) Features should capture all the significant and unique visual characteristics of the object of interest; (2) Features should be distinctive to the class of objects of interest and can adequately discriminate the object's appearance and shape from other objects; (3) Features should be robust enough, especially in the presence of background noise and changes in illumination conditions, camera viewpoint, and object's scale and orientation; and (4) Computation methods for these features