Automation in Construction 88 (2018) 23–30

Contents lists available at ScienceDirect

Automation in Construction

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

3D reconstruction of a concrete mixer truck for training object detectors

Hongjo Kim, Hyoungkwan Kim*

School of Civil and Environmental Engineering, Yonsei University, Seoul, Republic of Korea

ARTICLE INFO

Keywords:
3D reconstruction
Synthetic training data
Object detection
Computer vision
Construction site monitoring

ABSTRACT

Vision-based monitoring methods have been actively studied in the construction industry because they can be used to automatically generate information related to progress, productivity, and safety. Object detection is essentially used in such monitoring methods to infer jobsite context. However, as many classes of construction entities exist in a job site, large amounts of image data are required to train a detection algorithm to detect each class object in images. Although image data augmentation methods using 3D models were proposed, publicly available 3D models are limited to some construction object classes. Therefore, this study proposes a three-dimensional reconstruction method to generate the image data required for training object detectors. To use the generated synthetic images as training data, a histogram of oriented gradient (HOG) descriptor of a target object is obtained from these images. The descriptor is refined by a support vector machine to increase sensitivity to the target object in test images. The performance of the HOG-based object detector is evaluated using real images from ImageNet. The result shows that the proposed method can generate training data more effectively than existing manual data collection practices.

1. Introduction

Understanding the situation at a construction site is the most fundamental requirement for managing construction operations. When the current situation is understood, it is easy to identify existing or forthcoming problems affecting project cost, schedule, and worker safety. Such information as what objects exist, where they are, and when they appear in jobsites are inferred from the low-level data captured by monitoring systems. Thus, it is necessary to detect construction entities for jobsite monitoring. Among various types of sensing methods, computer vision technology has attracted considerable attention in construction site monitoring. Imaging technology enables object recognition in images and tracking of its behavior at low cost [1–5]. Typical applications of utilizing the technology in construction management (CM) are in safety management, productivity analysis, and progress monitoring [2,4–7].

Currently, the most common methods for recognizing object classes or extracting visual features are based on machine learning methods such as supervised learning [8,9]. Monitoring systems using supervised learning algorithms can identify objects in images, after being trained with image data containing labeled target objects. In general, a large amount of training data can enhance the performance of supervised learning algorithms. In the construction industry, however, the lack of publicly available image data makes it difficult to obtain a high performance of the system. If positive samples collected from a real construction site are used, a supervised learning algorithm for object detection can be overfitted (only works with a specific dataset in which target objects have very similar appearances with respect to the positive samples). This is because the algorithm did not capture general visual patterns of the target object class due to an insufficient number of the training samples, which lacks various appearances. Consequently, installing such monitoring system for a construction site involves manual collection of training data. This entails a time-consuming and labor-intensive task, because the system requires a considerable amount of training image data for each object class.

As an alternative to collecting training data, a method of generating synthetic image data using a three-dimensional (3D) model has been proposed [10]. However, if there is no existing 3D model for a target object, it is not an effective alternative because of the substantial time required to manually make the 3D model. While the 3D geometry of a target object can be directly collected by using laser scanners as reported in previous studies [3,11–14], reconstruction of the 3D geometry from 2D images taken by optical cameras is relatively difficult and requires a combination of multiple computer vision algorithms [15]. As numerous classes of construction objects exist in a construction site, preparing a 3D model for each object is also cumbersome. Another issue is that the efficacy of synthetic image data should be tested on a benchmark dataset having various appearances within an object class. Testing of such dataset can facilitate a reliable evaluation of synthetic images as training examples because the generalization capability of an

* Corresponding author.
E-mail addresses: hongjo@yonsei.ac.kr (H. Kim), hyoungkwan@yonsei.ac.kr (H. Kim).

https://doi.org/10.1016/j.autcon.2017.12.034
Received 16 January 2017; Received in revised form 9 November 2017; Accepted 27 December 2017
0926-5805/

algorithm can be measured on the test dataset unbiased to certain examples.

To address the abovementioned issues, this study proposes a 3D modeling method using two-dimensional (2D) video frames. The procedure is as follows. First, an unmanned aerial vehicle (UAV) is used to capture a video focusing on a target object in all directions. Second, the blurriness of video frames is calculated to filter out blurry frames to improve the quality of the image data. Third, a structure from motion (SfM) algorithm is used to estimate camera parameters on each image and to construct a 3D point cloud. Fourth, Poisson surface reconstruction algorithm is used to build a mesh surface on the point cloud. Fifth, the original texture of the object is mapped onto the mesh surface, as needed. Sixth, the reconstructed 3D model is projected to planes to generate 2D synthetic images in specified viewpoints. To demonstrate the efficacy of the proposed method, a concrete mixer truck is photographed and reconstructed in a 3D model. A HOG descriptor-based detector is trained using four synthetic images from the right-side views of the 3D model. The performance of the object detector is tested on 90 real images. The result shows that the proposed method can be an alternative to manual data collection. This study has been initiated based on two previous studies [16,17], and the contents have been strengthened and updated. The proposed system is illustrated in Fig. 1.

2. Related works

2.1. Vision-based CM applications and limitations

Recognizing construction objects in images enables automation of diverse CM applications that have conventionally been conducted by manual inspection. For safety management, Kim et al. [18] assessed a safety level for construction entities using fuzzy inference and computer vision-based monitoring. Seo et al. [19] presented a biomechanical analysis on lifting tasks using a vision-based capturing of worker's motion. Park et al. [20] proposed a detection method to check the wearing of safety helmets of workers. In this study, a worker is detected using a HOG template, and then the system judges whether the worker is wearing a safety helmet or not by searching the upper part of the detected bounding box. Chi and Caldas [21] presented a system that detects the violation of safety rules by earthwork equipment operators using a stereo vision camera. For productivity analysis, Bügler et al. [22] used photogrammetry and video analysis for estimating the earthwork productivity of dump trucks and excavators. Rezazadeh Azar et al. [23] estimated dirt-loading cycles by tracking the interaction of an excavator and a dump truck. Gong and Caldas [24] proposed a video interpretation method to investigate construction operation productivity using visual recognition and tracking of construction entities. Zou and Kim [25] calculated the idle time of an excavator by its centroid movement, segmenting the excavator in the color space of hue, saturation, and value. For progress monitoring, Dimitrov and Golparvar-Fard [26] proposed an appearance-based material classification method to monitor construction progress, using texture information of material classes. Kim et al. [27] updated a four-dimensional computer-aided design (CAD) model by comparing the as-planned model to the as-built structure in images. Golparvar-Fard et al. [28] visualized the current progress of construction sites by reconstructing a 3D point cloud from unordered images and comparing it with the existing building information model.

To automate diverse CM applications, the number of recognizable objects in construction sites should be increased. Kim et al. [29] proposed to increase the number of recognizable objects using the scene parsing method [30] that annotates each pixel of a query image with a class label. A total of 119 construction object classes could be detected by their method with a relatively small amount of training data. However, it was observed that a higher level of classification accuracy was achieved for those classes with a larger amount of training samples. The result implies that a large amount of training data is still required to obtain a stable performance of a monitoring system. This is a common phenomenon in studies that used machine learning algorithms for classification or detection [20,21,23,26,29]. However, previous studies lack studies to systematically collect training data. Moreover, it is necessary to validate detection models with an authorized benchmark dataset, in order to evaluate the generalization performance.

Previous studies have been conducted to develop methods for training object detectors using 3D models [31–34]. These studies have confirmed the potential of synthetic images taken from 3D models as a training data. In the construction domain, Soltani et al. [10] generated synthetic image data, composing the images of a 3D CAD model with real backgrounds. The study shows that synthetic images can be an effective training data for object detectors. Although the previous studies using synthetic images shows a capability to train object detectors, its use is limited to cases where 3D models already exist. To increase 3D model availability, we propose an automatic 3D modeling method from 2D images. Furthermore, we collect a challenging benchmark dataset and use it for evaluating the performance of object detectors trained with the synthetic data.

2.2. Multiview stereo and object detection algorithms

Reconstruct a 3D model from images needs multiview stereo algorithms, which can be classified into four categories: (1) voxel-based, (2) deformable polygonal mesh-based, (3) multiple depth map-based, and (4) patch-based algorithms [35,36]. Among the four classes of multiview stereo algorithms, patch-based algorithms are simple and effective to reconstruct a 3D model [35]. Because of these advantages, patch-based algorithms, such as SfM, have been widely used in the construction domain for progress monitoring [28,37,38], facility inspection [39–41], or energy performance monitoring [42–47]. This study also adopts a patch-based multiview stereo algorithm.

A patch-based multiview stereo method, SfM, used in this study builds a 3D model of a construction entity from 2D images. This method restores 3D information based on 2D image data [15,35,48,49]. First, the method reconstructs 3D geometric information by matching the same points across images. To attain this, a salient feature such as a scale-invariant feature transform (SIFT) descriptor has to be extracted from all images. The SIFT descriptor is a well-known salient feature because it is less sensitive to image scale, rotation, affine distortion, 3D viewpoint, noise, and illumination [50]. Next, an SfM algorithm estimates the intrinsic and extrinsic parameters of each image and

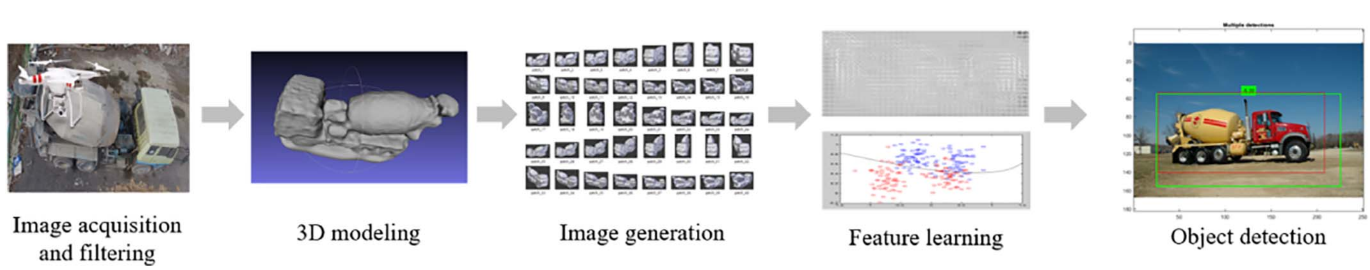

Fig. 1. The proposed system overview.