Automation in Construction 48 (2014) 1–10

Contents lists available at ScienceDirect

Automation in Construction

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

Image dataset development for measuring construction equipment recognition performance

Humaira Tajeen [1], Zhenhua Zhu *

Department of Building, Civil, and Environmental Engineering, Concordia University, Montreal H3G 1M8, Canada

ARTICLE INFO

Article history:
Received 21 August 2013
Received in revised form 31 May 2014
Accepted 15 July 2014
Available online 7 September 2014

Keywords:
Automation
Construction equipment
Dataset
Evaluation
Performance characteristics

ABSTRACT

The recognition of construction operational resources (equipment, workers, materials, etc.) has played an important role in achieving fully automated construction. So far, many object recognition methods have been developed in computer vision; however, they have been tested with a few categories of objects in natural scenes. Therefore, their performance on the recognition of construction operational resources is unclear, especially considering construction sites are typically dirty, disorderly, and cluttered. This paper proposes a standard dataset of construction site images to measure the construction equipment recognition performance of existing object recognition methods. Thousands of images have been collected and compiled, which cover 5 classes of construction equipment (excavator, loader, dozer, roller and backhoe). Each image has been annotated with the equipment type, location, orientation, occlusion, and labeling of equipment components (bucket, stick, boom etc.). The effectiveness of the dataset has been evaluated with two well-known object recognition methods in computer vision. The results show that the dataset could successfully identify the performance of these methods in terms of correctness, robustness, and speed of recognizing construction equipment.

1. Introduction

The construction industry has become one of the largest industrial sectors in Canada [1]. Similar to other industrial sectors, promoting automation in construction could be beneficial. For example, the automation of construction tasks could speed up construction processes, guarantee the consistency of construction operations, and perform construction jobs beyond human capabilities in size, weight, etc. [2]. This way, construction work could be done with the high productivity, little amount of rework, and reduced requirement for skilled workers. Meanwhile, the construction projects will more consistently be finished on time, within budget, and with high quality assured.

Considering these potential benefits, construction researchers and professionals have been working hard at promoting automation in construction, and a significant amount of automation work has been developed which analyze construction site images. High-resolution digital cameras have been increasingly adopted at construction sites due to their acceptable return on investment [3]. The time-lapse images collected by these cameras from construction sites do not only record the as-built progress of the projects under construction, but also capture the daily job site activities. Therefore, they could provide useful management information for construction engineers/managers to monitor and control sites remotely and dynamically.

In order to fully utilize construction site images in automating of construction work, one of the critical steps is the automatic recognition of various construction operational resources (e.g. equipment, workers, and materials). Once such construction operational resources are successfully recognized, many construction tasks can then be automated. However, the automatic recognition of construction resources under real construction site conditions is not easy. Typically, construction sites are characterized as dirty, untidy and cluttered with machines, tools, materials and debris. It is common that the resources on construction sites be only partially visible. All these characteristics make the recognition of on-site construction resources difficult.

So far, many object recognition methods have been developed, most of which were created by researchers in computer vision. These methods can be classified into three categories based on the recognition cues they have adopted. For example, Torralba et al. relied on the surface patches in images to recognize objects, and the recognition speed was further increased with the boosting technique [4]. Felzenszwalb et al. used histogram of oriented gradients (HOG) features, and developed a discriminatively-trained, part-based recognition method [5].

Typically, the performance of existing recognition methods is evaluated with several image datasets publically available, including the INRIA person dataset by Dalal and Triggs [6], the MIT-CSAIL dataset byTorralba et al. [4], the CALTECH dataset by Griffin et al. [7], the PASCAL VOC dataset [8], etc. To the authors' knowledge, these datasets only contain limited classes of objects in natural scenes, such as

* Corresponding author. Tel.: +1 514 848 2424x5948; fax: +1 514 848 7965.
E-mail addresses: humaira.tajeen@gmail.com (H. Tajeen), zhenhua.zhu@concordia.ca (Z. Zhu).
[1] Tel.: +1 514 519 0422.

http://dx.doi.org/10.1016/j.autcon.2014.07.006
0926-5805/© 2014 Published by Elsevier B.V.

pedestrians, human faces, bicycles, cars, etc. As a result, it is unknown whether they can be effectively used to evaluate the performance of existing object recognition methods for the recognition of on-site construction operation resources under real site conditions.

In order to address this issue, this paper outlines a new image dataset which includes typical construction equipment (excavator, loader, dozer, roller and backhoe) with a wide variety of sizes, poses, and camera viewpoints. Thousands of construction site images with significant illumination variations, and partial occlusions by debris, materials and other equipment have been collected and compiled. Also, a MATLAB-based annotation tool based on the work of Russel et al. [9] has been created to annotate the Equipment of Interest (EOI) in these images. The annotations include equipment type, location, viewing angle, occlusions, and labels of corresponding equipment components, such as bucket, stick, boom, cab, tracks, wheels, etc.

The database is expected to evaluate the construction equipment recognition performance of existing object recognition methods. The standard, unbiased, and extensive evaluation results from the use of the image dataset can help construction researchers and professionals select a better or more appropriate recognition method to automate construction tasks, including but not limited to productivity analysis, progress monitoring, etc. In order to evaluate the value of the image dataset, the database was tested with two well-known object recognition methods developed by Torralba et al. [4] and Felzenszwalb et al. [5]. The test results show that the image dataset developed in this paper can successfully identify the benefits and limitations of these methods when they are used to recognize different types of construction equipment on construction sites. Based on the benefits and limitations, recommendations can be made, based on whether the focus is placed on recognition correctness, robustness, speed or any combination of these. This way, construction researchers and professionals can select a better recognition method, if they need construction equipment recognition to automate their construction tasks. The dataset developed in this paper was tested on two object recognition methods, but it can be used to test other object recognition methods available.

2. Related work

2.1. Categories of 3D object recognition methods

Three-dimensional (3D) object recognition from images is considered as a challenging task. This is because the appearance of a 3D object can change significantly at different angles and poses to the camera. Also, the object can be partially occluded under heavily cluttered background, and experience different environmental lighting conditions [10]. So far, many object recognition methods have been developed, especially in the field of computer vision. Based on their recognition cues, the methods can be generally classified into three main categories: (1) geometry-based, (2) appearance-based, and (3) feature-based [10].

2.1.1. Geometry-based category

In geometry-based methods, object geometric primitives (e.g. lines, circles, and cylinders) are first selected to represent an object without other object properties such as color and texture [11]. These geometric primitives can be from 2D image contours [12], gradient response maps [13], or 3D synthetic Computer Aided Design (CAD) models [14,15]. After the extraction of the geometric primitives, a hierarchical organization of the primitives from multiple views is created. The object is then recognized if its geometry in the test image is similar to the geometric information contained in the hierarchical organization through template matching with bags of boundaries [12], ultra-wide baseline matching [14], etc.

These methods were supposed to be robust in the recognition of 3D objects, since the geometric primitives were invariant to viewpoint and illumination [16]. However, the test results showed that the reliable extraction of object geometric primitives could only be achieved under limited or controlled conditions (e.g. small degree of occlusions, background clutter, or lighting variations) [10]. In addition, it was difficult for the geometric primitives to capture the object deformations; and therefore, they were recently integrated in the context of local feature-based methods [17].

2.1.2. Appearance-based category

Appearance-based methods refer to those methods that rely on object color and texture as recognition cues, without the need for object-related geometric information. Under these methods, the recognition cues (color, texture, etc.) are first extracted with visual pattern recognition algorithms and represented as histograms [18]. They are then classified with the support of k-nearest neighbor, neural network with radial basis function, support vector machines, sparse network of Winnows, and others [10].

In general, these methods are simple and efficient, since they typically project the raw visual features to a lower dimensional feature space for the classification purpose [17]. However, they are less effective in handling object occlusions under cluttered backgrounds [10]. Also, it is time-consuming and labor intensive to get the training examples [10]. The object in the training samples must be manually segmented for the methods to learn its appearance characteristics [17].

2.1.3. Local feature-based category

Recently, local feature-based object recognition methods have gained popularity. The basic idea behind these methods is to represent an object with a set of local visual features. The object detection can then be done by matching these local features to the similar-looking ones in the test images [19]. The local visual features are typically invariant to scale, illumination and affine transformation. Currently, there are several well-known visual feature detectors and descriptors that have been created, including the Scale-Invariant Feature Transform (SIFT) [20], the Histogram of Oriented Gradients (HOG) [6], and the Speeded Up Robust Features (SURF) [21].

These methods have been recognized for their robustness due to the reliance on scale-, illumination-, and/or affine transformation- invariant visual features [11]. They do not always require the match of all object features for the successful recognition, which makes the detection applicable even when the object is partially occluded under the cluttered background [20]. Moreover, the training of these methods can be done automatically, since the object does not have to be fully segmented from the image background like the appearance-based methods require [11,17]. To date the local feature-based methods have been preliminarily applied to a number of tasks, not only in computer vision, but also in the construction field for the recognition of trucks [22], construction workers [23], etc. However, empirical evidence has shown that most recognition methods are fragile and unable to generalize the recognition of objects in new environments [17].

2.2. Datasets available for object recognition performance evaluation

There has been tremendous progress made towards object recognition; however, existing recognition methods are still not perfect. They are sensitive to large illumination variations and heavy occlusions [10]. In order to evaluate the performance of existing object recognition methods, several datasets have been developed. However, they include a large number of images which cover limited categories of objects in natural settings. For example, the dataset developed at MIT contains bicycle, bottle, apple, bookshelf, car, chair, desk, sofa, building, door, window, etc. [9]. The dataset at UIUC only includes cars with side views [24]. The INRIA dataset was created as a part of research work in human detection, which contains images of people in upright positions [6]. The PASCAL VOC dataset contains twenty visual object classes: person, bird, cat, cow, dog, horse, sheep, airplane, bicycle, boat, bus, car, motorbike, train, bottle, chair, dining table, potted plant, sofa, and TV/monitor [8]. The CALTECH-101 dataset contains 101 classes of objects including