Automation in Construction 62 (2016) 14–23

Contents lists available at ScienceDirect

Automation in Construction

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

Automated annotation for visual recognition of construction resources using synthetic images

Mohammad Mostafa Soltani [a], Zhenhua Zhu [b], Amin Hammad [c,*]

[a] Building, Civil and Environmental Engineering, Concordia University, 1455 de Maisonneuve Blvd. West, EV-6.139, Montreal, QC H3G 1M8, Canada
[b] Building, Civil and Environmental Engineering, Concordia University, 1455 de Maisonneuve Blvd. West, EV-6.237, Montreal, QC H3G 1M8, Canada
[c] Concordia Institute for Information Systems Engineering, Concordia University, 1515 Ste-Catherine Street West, EV7.634, Montreal, QC H3G 2W1, Canada

ARTICLE INFO

Article history:
Received 12 February 2015
Received in revised form 13 October 2015
Accepted 16 October 2015
Available online 18 November 2015

Keywords:
Object recognition
Construction equipment
Synthetic images
Auto-annotation
Auto negative image sampler

ABSTRACT

The recognition of construction equipment is always necessary and important to monitor the progress and the safety of a construction project. Recently, the potentials of computer vision (CV) techniques have been investigated to facilitate the current equipment recognition method. However, the process of manually collecting and annotating a large image dataset of different equipment is one of the most time-consuming tasks that may delay the application of the CV techniques for construction equipment recognition. Moreover, collecting effective negative samples brings more difficulties for training the object detectors. This research aims to introduce an automated method for creating and annotating synthetic images of construction equipment while significantly reducing the required time. The synthetic images of the equipment are created from the three-dimensional (3D) models of construction machines combined with various background images taken from construction sites. The location of the equipment in the images is known since that equipment is the only object over the single-color background. This location can be extracted by applying segmentation techniques and then used for the annotation purpose. Furthermore, an automated negative image sampler is introduced in this paper to automatically generate many negative samples with different sizes out of one general image of a construction site in a way that the samples do not include the target object. The test results show that the proposed method is able to reduce the required time for annotating the images in comparison with traditional annotation methods while improving the detection accuracy.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The application of computer vision (CV) is growing within the construction sector. Monitoring the productivity and the progress of the project, and tracking the equipment for safety purposes are the targets of many research studies using CV. Object recognition is the first and fundamental step toward the above-mentioned goals. For example, estimating the productivity requires to follow the fleet of the resources on the construction site. For this reason, the resources have to be identified and positioned within the images taken from the site. The main prerequisite for object recognition is to train the CV object recognition algorithms so that they will be able to find similar objects in the images. Usually, the training for object recognition is a time-consuming and sensitive task which has a direct effect on the accuracy of the object recognition results. Furthermore, the blind training of a recognition model with a large number of samples without a solid structure can create other problems such as fewer true positive detections or more false positive detections even after spending a lot of time in training with those samples. Studying the process of training the vision-based object detectors shows that annotating the region of interest (ROI) within each image needs a lot of time and effort [23,58].

This paper elaborates on the possibility of using a three-dimensional (3D) model of construction equipment (e.g. excavators) instead of the images of the equipment to automatically generate annotations for object recognition training. The 3D model can be automatically integrated with different backgrounds of the construction sites to improve the training quality and reduce the training time. The resulting high-quality training with the well-structured annotation could significantly improve the object recognition. Moreover, this method helps the industry to have access to numerous image datasets for different equipment from different manufactures and to customize the supervised object detectors for each target object. Also, extracting more information from the images, such as the position and the orientation of the equipment parts, helps in tracking the pose of the equipment for the safety and productivity evaluation purposes.

In addition, collecting an effective negative dataset is another challenge for training a detector model. The importance of negative samples is not less than positive samples since a robust detector not only needs to detect the target object correctly but also it has to reject the false

* Corresponding author. Tel.: +1 514 848 2424x5800; fax: +1 514 848 3171.
E-mail addresses: mo_solta@encs.concordia.ca (M.M. Soltani), zhzhu@bcee.concordia.ca (Z. Zhu), hammad@ciise.concordia.ca (A. Hammad).

http://dx.doi.org/10.1016/j.autcon.2015.10.002
0926-5805/© 2015 Elsevier B.V. All rights reserved.

detection. Generating negative samples related to construction projects can be automated by randomly cropping few images taken from construction sites.

The current paper is an extension of a short conference paper representing the preliminary results of the research [45]. This paper aims to achieve the following objectives: (1) to automatically annotate the synthetic images created by overlaying the 3D model of the construction equipment on background images of the construction site, (2) to automatically generate effective negative samples for training the detectors, and (3) to apply sensitivity analyses to determine the effect of the number of step angles for each detector and the effect of the covered range of angles of view associated with each detector. Furthermore, the proposed method is evaluated by comparing the results of object recognition using the proposed method and the results achieved using manual annotation of the real objects.

2. Literature review

Supporting the proposed method in this research requires studying the applications of CV in construction, the methods of object recognition, and the current annotation and labeling approaches. In this section, the details about the current state of object recognition in construction sites are investigated. Also, the current development in the areas of image annotation and segmentation is studied.

2.1. Construction equipment and workers recognition

An object can be detected and categorized through its shape, color, and/or motion characteristics [37]. The following studies take advantage of each of these characteristics (or a combination of them) to recognize the workers and/or equipment in construction projects.

Zou and Kim [60] in an early attempt that studied the application of hue, saturation, and value (HSV) color space to estimate the idle time of the hydraulic excavators. Comparing red, green, and blue (RGB) color space with HSV shows that RGB color space is easy to understand, but it brings many difficulties to differentiate the object of interest from the background (especially when the light conditions change from bright to dark) since the histogram of RGB values of the target object usually overlap with the histogram of its background. On the other hand, the characteristics of hue change less by changing the light conditions in open fields since the HSV color space relies on the dominant wavelength of the perceived color. One of the assumptions in [60] is selecting the object of interest in the original image manually by the user which limits the application of the proposed method.

Weerasinghe and Ruwanpura [57] proposed detecting the workers through their hardhats. The algorithm searches for the hardhats with a known color (e.g. yellow). However, this method is not reliable for complex construction sites with a cluttered background.

Template matching is another way to detect the appearance of an entity [10], which can be applied using object eigen-image [35,53], an active shape or appearance model [11,12,33], and a bag of words or textons ([13,24,43]). Brilakis et al. [8] used a semantic texton forests (SFTs) method which learns the appearance features of the object and context information [43].

Another research done by Chi and Caldas [9] introduces the background subtraction method developed first by Li et al. [27] for detecting the construction resources. This method detects the objects by finding and removing the static pixels (background) out of the moving pixels (foreground). Knowing the foreground pixels can be used to find the moving object, but it does not provide any further information about the object (whether it is a worker, a backhoe, or a truck). Therefore, the foreground pixels are passed through a pre-trained object classifier (i.e., a Bayes classifier or a neural network) and the classifier determines the type of the moving object. As a limitation, this method needs the object to be moving in the subsequent frames to be detected; but if the object is stationary, it is not recognized.

Haar-like features supported by adaptive boosting algorithm [17,54], HOG features coupled with a support vector machine (SVM) developed by [14], color histogram [48], and Eigen-images including color and shape information, can be adapted for construction resources' recognition [37]. Park and Brilakis [38] proposed applying the background subtraction method coupled with Haar-cascade features. Moreover, they used HSV color space to minimize the false detections by considering the colors of the objects. In another research, Parl and Brilakis [38] investigated the combination of background subtraction method, HOG features with SVM classifier, and color histogram with k-NN (k Nearest Neighbors) in subsequent levels to recognize the worker at construction sites. However, both methods are limited to the moving objects.

Azar and McCabe [4] studied the recognition result by using HOG and Haar-Like features on the static images (Haar–HOG). They trained eight detectors to cover all views around the equipment. Moreover, they applied Haar detector in combination with HOG detector on the video of the site. In another scenario, the foreground object of the video was detected and then the HOG detector searches for the target object within the foreground pixels (Blob–HOG). The results show that Haar–HOG performs slightly better than Blob–HOG, while the Blob–HOG is less intensive in terms of computation.

Memarzadeh et al. [34] proposed using HOG features in association with histogram of color (HOC) and their results show that HOG–HOC slightly outperformed HOG. In the research done by Tajeen and Zhu [49], two methods previously developed by Torralba et al. [51,52] and by Felzenszwalb et al. [16] were compared in the construction environment. The idea that Torralba et al. [51,52] presented is to use the shared patches of different classes through a machine learning process. This concept helped to run the classifier faster while it requires less data to train since it uses the shared data across the classes. Felzenszwalb et al. [16] proposed the discriminatively trained part-based model, which not only uses HOG features of the whole deformable object but also considers HOG features of each attached subpart of that object. It also applies a latent SVM to formulate the relations between the features of the object and its parts. The comparison results provided by Tajeen and Zhu [49] show that the part-based model performs better and is more robust to occlusions, while the sharing features method performs faster.

The limitation of some of the mentioned methods from the spent time point of view is explained in Section 4.1 by comparing the estimated time for manually annotating the image datasets with the proposed method in this research.

2.2. Annotation tools

Assuming the use of supervised object detectors, it is necessary to train a descriptor for recognizing a specific object within an image. The training of the object detectors starts by annotating many training samples which contain that object. Typically, the samples are observed one by one and the target object in each sample is identified with a rectangular bounding box to indicate the regions that are occupied by the object. The coordinates of the bounding box are considered as the region of interest (ROI) during the training phase. In the following, different approaches for image annotation, which suffer from the long time required for manual annotation, are studied.

There are a few toolboxes available for annotating and labeling the images that can be used for training purpose. Von Ahn and Dabbish [55] and Von Ahn et al. [56] proposed an online computer game named Peekaboom for labeling images through an interactive system. The game is arranged so that the user determines the contents of images by choosing meaningful labels for them.

LabelMe is one of the well-known tools in this area [40]. It is a web-based tool which has the ability to be extended to automatically enhance object labels with WordNet [15], discover object parts, recover a depth ordering of objects in a scene, and increase the number of labels using minimal user supervision. A semi-automatic labeling tool in this

Download English Version:

https://daneshyari.com/en/article/246230

Download Persian Version:

https://daneshyari.com/article/246230

Daneshyari.com