Automation in Construction 81 (2017) 161–171

Contents lists available at ScienceDirect

Automation in Construction

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

Integrated detection and tracking of workforce and equipment from construction jobsite videos

Zhenhua Zhu*, Xiaoning Ren, Zhi Chen

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

ARTICLE INFO

Keywords:
Visual detection
Visual tracking
Imaging techniques
Automatic identification systems

ABSTRACT

High definition (HD) video cameras have been used to record daily activities at construction jobsites into videos. These videos contain rich workforce and equipment information for site engineers and project managers to analyze construction productivity, monitor construction progress, inspect jobsite safety, etc. However, it is difficult to automatically retrieve such information from the videos, since existing methods for the detection of construction workforce and equipment could not reach high precision and recall at the same time. This paper presents a novel framework that integrates the visual tracking into the detection of construction workforce and equipment. The integration significantly improves the recall and meanwhile maintains high precision. The proposed framework has been tested in real construction jobsites. Although it does not process the jobsite videos in real time yet, the test results showed that the recall for the detection of construction workforce and equipment was improved by more than 30–50%, while maintain the precision at the same level.

1. Introduction

High definition (HD) video cameras have been recently installed at construction jobsites, considering their tangible returns on investment [1]. The cameras are expected to capture detailed daily jobsite activities of construction workforce and equipment into videos. Suppose one camera takes videos at one frame per second (FPS). It will produce large sized videos by the end of each working day. These videos document construction events and processes. They facilitate the work of field engineers and/or project managers on recording the working hours of the workforce and equipment at construction jobsites, analyzing their construction productivities [2], and even monitoring their potential jobsite safety issues [3].

Currently, it is difficult to automatically retrieve the detailed workforce and equipment information from the jobsite videos. The information includes, but is not limited to, 1) the equipment types and worker trades, 2) the number of the equipment and workers onsite; 3) the work zones where the equipment and workers are; and 4) the working states of the equipment and workers. If such information contained in the videos has to be manually retrieved by exploring one video to another, a lot of manual efforts are required. These manual efforts add the workloads of site engineers and/or managers and indirectly increase project overheads to construction contractors. Therefore, researchers in the construction field have been investigating novel solutions that rely on image processing and computer vision techniques to automatically retrieve the workforce and equipment information from the daily construction jobsite videos.

Most of existing methods created by the researchers are based on the detection of the workforce and equipment from the jobsite videos [4]. The detection is always a fundamental step in the process for the retrieval of the workforce and equipment information [5]. However, the sole reliance on the detection produced a significant challenge, considering it is always difficult to guarantee high precision and recall for the retrieval of the workforce and equipment information at the same time [6]. Typically, when more detections are produced by the methods in the jobsite videos, it is possible to locate more workers and equipment items (i.e. high recall), but meanwhile more wrong detections are also introduced (i.e. low precision). In contrast, when fewer detections are produced, most of those detections could be correct (i.e. high precision) but not many workers and equipment items could be located (i.e. low recall).

The main objective of this paper is to achieve both high precision and recall at the same time by effectively combining the visual detection and tracking of the workforce and equipment in construction jobsite videos. In order to achieve this objective, a novel vision-based framework is presented here, which integrates the visual tracking into the detection of construction workforce and equipment. Both detection and tracking modules work together to locate the workers and equipment in the jobsite videos. The missed detections of the workers and equipment in the videos could be tracked, while the failure of the

* Corresponding author.
E-mail addresses: zhenhua.zhu@concordia.ca (Z. Zhu), ryanren528@gmail.com (X. Ren), zhi.chen@concordia.ca (Z. Chen).

http://dx.doi.org/10.1016/j.autcon.2017.05.005
Received 6 June 2015; Received in revised form 28 April 2017; Accepted 9 May 2017
Available online 20 June 2017
0926-5805/

tracking could be corrected by the detections. This way, the detection and tracking modules are complementary each other to identify the workforce and equipment from the jobsite videos with high precision and recall.

The proposed framework has been implemented in the Matlab R2014b environments. Its effectiveness has been tested in real construction sites. Although the framework does not process the jobsite videos in real time yet, the test results showed that the recall for the detection of construction workforce and equipment was improved by more than 50%, compared with the sole reliance on the detection. Meanwhile, the precision was still maintained around 90%. It is also worth noting that any of existing workforce and/or equipment detectors created by other researchers could be adopted in the framework. The novelty of this research does not lie in creating new workforce and equipment detectors to improve the detection performance. Instead, it is to find an effective way to integrate the tracking into the workforce and equipment detection procedure.

2. Related work

This section mainly present existing research studies related to the detection of construction workers and equipment in construction jobsite videos. It is then followed by the descriptions of the methods available for visually tracking construction workers and equipment. Both are what the proposed framework plans to build on.

2.1. Visual detection of construction workforce and equipment

The visual detection of construction workers and equipment is always a fundamental step in the process of using jobsite videos to automate construction engineering and management tasks. The detection could help engineers and managers retrieve important construction activity information to facilitate their analyses and decision makings at the jobsites. So far, there are several research studies that have been performed to investigate the potential of visually detecting construction workers and equipment in the jobsite videos. In those studies, background subtraction is one of the most prevalent methods for the purpose of detecting construction workers and equipment in motion [7]. Typically, it is used to determine and extract the motion pixels in the video streams. Then, the motion pixels are grouped, so that the regions of moving workers and equipment could be identified as foreground and other static regions are eliminated as background. One example of using the background subtraction to detect moving construction workers and equipment could be found in the work of Chi and Caldas [8], where the background subtraction algorithms were tested at equipment-intensive construction sites. Gong and Caldas also evaluated three types of background subtraction algorithms (i.e. Mixtures of Gaussian, Codebook based, and Bayesian model based methods) [5]. All these three algorithms were tested to detect construction workers and equipment; and the results showed that the Mixtures of Gaussian method was not appropriate and the Codebook based algorithm was desirable [5]. However, one major limitation for the detection of construction workers and equipment with the background subtraction lies in its ability to only identify moving workers and equipment in construction scenarios. Therefore, when workers are standing or equipment is idle at construction jobsites, it is difficult for the background subtraction algorithms to locate them. Also, the background subtraction algorithms could only extract the regions of the moving workers and equipment from the video background [3]. They fail to classify these regions and identify which one represents a worker and which one belongs to equipment as well as what type of the equipment.

Also, the detection methods that rely on visual features (e.g. color and shape) have been proposed and tested in construction scenarios. Histogram of oriented gradients (HOG) and Haar-like features are two popular shape-based features that could be used to detect construction workers and equipment [2,7]. For example, Rezazadeh Azar and McCabe presented two detection methods, Haar-HOG and Blob-HOG to detect dump trucks for the purpose of measuring the productivity of earthmoving operations [2]. Moreover, in the detection module of their server custom interaction tracker (SCIT), the HOG features under the support of Graphic Processing Unit (GPU) computation were adopted [9]. A similar research study could be found in the work of Memarzadeh et al., where they combined the HOG features a new multiple binary Support Vector Machine (SVM) classifier to automatically detect and distinguish workers and equipment in the video streams [10].

Compared with the detection methods based on shape features, the detection methods that rely on color features are expected to be simple and effective, especially when the construction workers and/or equipment of interest to detect are uniquely colored. Although the results indicated that the color-based detection methods have the possibility to detect construction equipment, the sole reliance on the color information could easily make the methods vulnerable to illumination changes [11]. Therefore, the methods solely based on color information as detection cues might fail, considering that construction activities could be conducted from day to night.

Moreover, the shape and color features could be combined as detection cues to improve the detection performance. For example, Memarzadeh et al. developed a novel detection methods based on HOG and Colors (HOG + C) to detect workers and equipment in site videos [12]. Park and Brilakis presented the combination of the HOG and the histogram of HSV color values to automatically detect construction workers [13]. The HOG features are used to detect and localize personnel at a construction jobsite first; and then the histogram of the HSV color values is utilized to identify those construction workers by checking whether the personnel are wearing safety vests or not. A more comprehensive combination for the detection of construction workers and equipment could be found in the work of Park and Brilakis, where they combined not only shape and color features, but also motion features as detection cues [7].

In addition to the research studies mentioned above, Azar et al. tried to detect construction equipment and estimated its pose using the low-cost visual markers attached on the equipment [14,15]. They concluded that the detection precision with the visual markers almost reached 100% [14]. Liu and Golparvar-Fard introduced the intelligence of the crowd for interpreting jobsite videos to analyze construction activities from jobsite video streams [16]. Yuan and Cai presented a kinematic key nodes model for dynamically detecting and locating movable objects at construction sites [17].

Moreover, the recent research studies in the computer vision community showed the significant improvement on the object detection accuracy. This is especially true for the pedestrian detection, which might help the researchers in the construction field create appropriate methods for the detection of construction objects (e.g. workers). One example is the application of convolutional neural networks (CNNs). Girshick et al. relied high-capacity CNNS to localize and segment objects [18]. However, Dollár et al. concluded that the detection performance still has much room for improvement after they evaluated the state of the art pedestrian detection methods [19]. Also, their effectiveness on the detection of construction objects at real construction jobsites has not been verified yet.

2.2. Visual tracking of construction onsite resources

The visual detection is mainly used to identify and locate the workers and equipment in jobsite videos, while the visual tracking could be used to follow the movement of the workers and equipment and determine their corresponding trajectories. The tracking of construction workers provides engineers and managers with timely feedbacks when being used to assess the labor productivity of the workers and monitor their safety-related issues at construction jobsites. Considering these potential benefits, Teizer and Vela compared and evaluated four existing visual tracking algorithms (i.e. Mean-shift,