Automation in Construction 48 (2014) 74–87

Contents lists available at ScienceDirect

Automation in Construction

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

Vision-based workface assessment using depth images for activity analysis of interior construction operations

Ardalan Khosrowpour [a,1], Juan Carlos Niebles [b,2], Mani Golparvar-Fard [c,d,*]

[a] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States
[b] Electrical and Electronic Engineering Department, Universidad del Norte, Barranquilla, Colombia
[c] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N. Mathews Ave., Urbana, IL 61801, United States
[d] Department of Computer Science, University of Illinois at Urbana-Champaign, 205 N. Mathews Ave., Urbana, IL 61801, United States

ARTICLE INFO

Article history:
Received 13 March 2014
Received in revised form 18 June 2014
Accepted 19 August 2014
Available online 17 September 2014

Keywords:
Activity analysis
Workface assessment
RGB-D cameras
Hidden Markov Model

ABSTRACT

Workface assessment – the process of determining the overall activity rates of onsite construction workers throughout a day – typically involves manual visual observations which are time-consuming and labor-intensive. To minimize subjectivity and the time required for conducting detailed assessments, and allowing managers to spend their time on the more important task of assessing and implementing improvements, we propose a new inexpensive vision-based method using RGB-D sensors that is applicable to interior construction operations. This is a particularly challenging task as construction activities have a large range of intra-class variability including varying sequences of body posture and time-spent on each individual activity. The skeleton extraction algorithms from RGB-D sequences produce noisy outputs when workers interact with tools or when there is a significant body occlusion within the camera's field-of-view. Existing vision-based methods are also limited as they can primarily classify "atomic" activities from RGB-D sequences involving one worker conducting a single activity. To address these limitations, our method includes three components: 1) an algorithm for detecting, tracking, and extracting body skeleton features from depth images; 2) a discriminative bag-of-poses activity classifier for classifying single visual activities from a given body skeleton sequence; and 3) a Hidden Markov Model to represent emission probabilities in the form of a statistical distribution of single activity classifiers. For training and testing purposes, we introduce a new dataset of eleven RGB-D sequences for interior drywall construction operations involving three actual construction workers conducting eight different activities in various interior locations. Our results with an average accuracy of 76% on the testing dataset show the promise of vision-based methods using RGB-D sequences for facilitating the activity analysis workface assessment.

1. Introduction

Several recent research studies have shown the feasibility of construction activity analysis and its positive correlation with improved direct-work rates of the workers and also the overall jobsite productivity [1–6]. According to the Construction Industry Institute (CII) [2,4], successful implementation of activity analysis involves two key steps: (1) continuous workface assessment, and (2) planning and implementing improvements. As the first step, workface assessment – the process of determining the overall activity rates of onsite construction workers throughout a day – involves an observer walking along randomly selected pre-defined routes, and characterizing the activity of each observed worker [2]. Nevertheless, visual observation at high level of confidence is constrained by the high cost associated with performing manual data collection, the risk of interfering the activities under observation, and the tendency to produce inaccurate data [7,8]. For example, obtaining 95% confidence in workface assessment – considering the constraints above – requires a minimum of 5100 random observations for a 10 hour working shift regardless of the worker population, activity type, or the job size [9]. Consequently to avoid the over-productiveness phenomenon caused by close surveillance and observation of workers – the Hawthorne Effect – distance limit instructions for manual visual observations are proposed [9]. Initiating random routes and times, obeying standard distance limits to the workers, and instantaneous task-level data collection on entire job-site are some of the other key issues to be considered. Manual implementation of these tasks especially for several ongoing operations on a jobsite can significantly challenge frequent implementation of workface assessment, which is a necessary step before improvement can be planned and implemented [2,4,10].

To address current limitations, a large body of research in the past few years has focused on methods that can automate the workface

* Corresponding author. Tel.: +1 217 300 5226; fax: +1 217 265 8039.
E-mail addresses: khosrow2@illinois.edu (A. Khosrowpour), njuan@uninorte.edu.co (J.C. Niebles), mgolpar@illinois.edu (M. Golparvar-Fard).
[1] Tel.: +1 217 402 4847; fax: +1 217 265 8039.
[2] Tel.:+57 5 350 9509x3439.

http://dx.doi.org/10.1016/j.autcon.2014.08.003
0926-5805/

assessment process. These methods range from application of sensors such as Ultra Wide Band (UWB) systems [11–13], Radio Frequency Identification (RFID) tags [14], and Global Positioning Systems (GPS) [15,16] to computer vision methods using video cameras [17–19,71]. Several existing methods that build on top of the non-visual sensors mainly track the locations of the workers. Without interpreting the activities of the workers and purely based on location information – which is the basis of the majority of the state-of-the-art methods – deriving workface assessment data is challenging. For example for interior drywall activities, distinguishing between idle time, picking up a gypsum board, and measuring and cutting purely based on location data will be very difficult as during these activities the location of the worker would not necessarily change. Because these solutions do not yet perform accurately and do not produce a detailed feedback, the construction industry is still reluctant to adopt these automated solutions.

To address the limitations of location-based activity recognition, Joshua and Varghese [20–23] proposed an accelerometer-based method that has the capability of recognizing various activities based on movement of the body skeleton. Their method was tested for bricklaying operations at the task-level resolution and promising results have been reported. Using prior knowledge about activity locations on the jobsite, Cheng et al. [7] proposed an activity analysis method based on both location and body posture of the workers by integrating UWB – for location tracking – and commercially-available Physiological Status Monitors (PSM) with a wearable 3-axial thoracic accelerometer to derive body posture data. This method uses a single body posture and location to model and infer each activity. Still distinguishing between two activities that have the same location and body pose, for example idle time and measuring dimensions of a gypsum board, would be challenging.

Our method is different from prior research, as we choose to use inexpensive RGB-D sensors (<$150) that can provide confidentiality in the data collection, and can detect and track body skeleton of up to six workers simultaneously and in real-time. Confidentiality here means that the identity of the workers remain unknown as we only track their body skeleton. To generalize the applicability of our method, we do not assume any prior knowledge about expected activities in certain locations on the jobsite. Also rather than directly interpreting location and single body posture to derive activities as in [7], we propose histograms of body posture from RGB-D sequences to capture tabulated frequencies of a large number of key body postures for construction activities and use learning methods to train and infer these activities in a principled way. Our method includes: 1) an algorithm for detecting, tracking, and extracting body skeleton features from depth images captured using the RGB-D sensors; 2) a discriminative bag-of-pose activity classifier trained using multiple Support Vector Machines for classifying single visual activities from a given body skeleton sequence; and 3) a Hidden Markov Model (HMM) with Kernel Density Estimation (KDE) to represent emission probabilities in the form of a statistical distribution of single activity classifiers. For training and testing purposes, we introduce a new dataset of eleven RGB-D sequences for interior drywall operations involving three actual construction workers conducting eight different activities in various locations. Instead of manually collecting and analyzing workface data, the proposed method allows project managers to spend their time on correctly interpreting the results which is key to increasing productive activities in construction [6] and according to [24–27] requires more attention because conditions may differ from one project to another. In the following, we review the related work on vision-based methods.

2. Related work on vision-based methods

The advent of high-resolution video cameras, high storage databases, and availability of Internet over the past few years has transformed the ongoing construction operations' documentation methods. Today, it is common for owners and contractors to have cameras continuously monitoring their onsite construction activities. Building on the state-of-the-art algorithms in computer vision and leveraging these existing cameras, several methods have been proposed that focus on detecting construction workers and equipment [28–31], tracking their location in 2D and 3D [32,33], recognizing their activities based on their locations [34], or classifying atomic activities from videos containing a worker or equipment performing a single activity [35–37]. Teizer and Vela reviewed and compared existing computer vision tracking methods based on RGB images and highlighted the challenges of workface interaction such as occlusion, visual clutter, and photometric visual variability on construction sites. In addition to tracking worker location, activity recognition using video cameras has been a research area that has received attention over the past few years. To distinguish from construction activity analysis, by *worker activity recognition*, we mean detecting and documenting activities that are conducted by workers as part of the workface assessment process. Peddi et al. [17] proposed a method which classifies worker activity into three main categories – effective, ineffective, and contributory – based on a pose detection and tracking algorithms. Golparvar-Fard et al. [35] presented an algorithm that learns the distributions of spatio-temporal features and individual activity categories for earthmoving equipment using a multi-class Support Vector Machine (SVM) learning/inference model.

Over the last two to three years, the advent of depth sensors such as Microsoft Kinect, PrimeSense Carmine, and Time of Flight (TOF) cameras, has significantly facilitated the task of "detecting and tracking people". These sensors facilitate detecting, tracking body skeleton, and

Table 1
Summary of limitations of different technologies for automated workface assessment. For comparison of camera-based methods, the reader is encouraged to look into [28,35].

		State-of-the-art methods	RGB-D based solution
Technical features and capabilities	Detect and track individual workers	Yes	Yes
	Detect and track groups of workers	Yes. Challenges are (1) individual workers need to be tagged; (2) economy of scale as tracking tags for all permanent and temporary workers need to be supplied and maintained.	Yes. The sensor needs maintenance.
	Different environmental conditions	Limitations: (1) Needs access to GPS, UWB, and RFID tag readers; (2) coverage range varies based on the availability of the supporting location tracking infrastructure	Limitations: (1) can be used where there is no direct sunlight (primarily indoors); (2) limited by field-of-view and operating distance (<5 m with current sensors). Multiple RGB-D sensors can be used but interferences need to be managed.
	Recognize body posture	No. This likely requires several accelerometers to be attached to body.	Yes; in real-time, however occlusions and tool-interactions can affect the accuracy of detecting body postures.
	Derive activities from joint representation of location and body postures	(1) Without several accelerometers, the sequence of body postures may not be easily derived. (2) Deriving worker activities from such sequences is not explored.	Yes.
	Scalability of the solution	Requires coverage of location tracking infrastructure and one tag per individual worker.	Requires one to two sensors (each capable of covering up to 6 workers) for each work area.