Advanced Engineering Informatics xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Computer vision techniques for construction safety and health monitoring ☆

JoonOh Seo [a], SangUk Han [b], SangHyun Lee [a,*], Hyungkwan Kim [c]

[a] Tishman Construction Management Program, Dept. of Civil and Environmental Engineering, University of Michigan, 2350 Hayward St., Ann Arbor, MI 48109, USA
[b] Dept. of Civil and Environmental Engineering, University of Alberta, 9105 116th St., Edmonton, Alberta T6G 2W2, Canada
[c] School of Civil and Environmental Engineering, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Republic of Korea

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form 2 January 2015
Accepted 3 February 2015
Available online xxxx

Keywords:
Construction safety and health
Computer vision
Monitoring

ABSTRACT

For construction safety and health, continuous monitoring of unsafe conditions and action is essential in order to eliminate potential hazards in a timely manner. As a robust and automated means of field observation, computer vision techniques have been applied for the extraction of safety related information from site images and videos, and regarded as effective solutions complementary to current time-consuming and unreliable manual observational practices. Although some research efforts have been directed toward computer vision-based safety and health monitoring, its application in real practice remains premature due to a number of technical issues and research challenges in terms of reliability, accuracy, and applicability. This paper thus reviews previous attempts in construction applications from both technical and practical perspectives in order to understand the current status of computer vision techniques, which in turn suggests the direction of future research in the field of computer vision-based safety and health monitoring. Specifically, this paper categorizes previous studies into three groups—object detection, object tracking, and action recognition—based on types of information required to evaluate unsafe conditions and acts. The results demonstrate that major research challenges include comprehensive scene understanding, varying tracking accuracy by camera position, and action recognition of multiple equipment and workers. In addition, we identified several practical issues including a lack of task-specific and quantifiable metrics to evaluate the extracted information in safety context, technical obstacles due to dynamic conditions at construction sites and privacy issues. These challenges indicate a need for further research in these areas. Accordingly, this paper provides researchers insights into advancing knowledge and techniques for computer vision-based safety and health monitoring, and offers fresh opportunities and considerations to practitioners in understanding and adopting the techniques.

1. Introduction

Due to hazardous working environments at construction sites, workers frequently face potential safety and health risks throughout the construction process. Even though the construction sector constitutes about 5% of the workforce in the U.S., fatal injuries in construction account for about 18% of all occupational deaths [1]. In addition, the incident rate for nonfatal occupational injuries and illness in construction is 30% higher than average industries [2]. These statistics obviously show an immediate need to reduce the prevalence of fatal and non-fatal injuries in construction. To address this issue, previous research has examined accident occurrence mechanisms to understand the causes of accidents, and the process and conditions leading to accidents. Notably, prior work demonstrates that nearly all such injuries are highly preventable by reducing or eliminating exposures that may contribute to detrimental safety and health effects to construction workers [3,4].

In the accident causation model [5], unsafe conditions and unsafe acts are considered the two direct causes of accidents. Monitoring unsafe conditions and acts in the construction process thus plays a key role in determining and taking prompt corrective actions to prevent resulting safety and health issues by eliminating them in the causal process. In practice, site observations and inspections are commonly used techniques to evaluate the risk involved in ongoing works and existing site conditions in construction [6]. However, observational methods are costly and time-consuming because they require human manual observations

☆ Handled by W.O. O'Brien.

* Corresponding author. Tel.: +1 734 764 9420; fax: +1 734 764 4292.
E-mail addresses: junoseo@umich.edu (J. Seo), sanguk@ualberta.ca (S. Han), shdpm@umich.edu (S. Lee), youngkwan@yonsei.ac.kr (H. Kim).

http://dx.doi.org/10.1016/j.aei.2015.02.001
1474-0346/

and documentations by supervisors or safety personnel [7]. In addition, manual observation suffers from the limitations of missing and inaccurate information in a timely fashion [8]. These limitations become more significant in construction projects because worker environments continuously change over time and skilled supervisory manpower is not always present at sites, thereby making it challenging to implement manual observation processes in daily practice [9].

Recently, computer vision has drawn attention because it can be used for the automated and continuous monitoring at construction sites. Computer vision can provide a rich set of information (e.g., locations and behaviors of project entities, and site conditions) about a construction scene by taking images or videos, which facilitates the understanding of the complex construction tasks rapidly, accurately, and comprehensively. These advances bring the operational and technical advantages over other types of sensing techniques (e.g., RFID, GPS and UWB) that require installation of sensors to all of project entities to be monitored and provide limited information such as location data, providing an opportunity to complement them [10,13,23,44,45,79]. Accordingly, computer vision has been applied to various areas in construction such as progress monitoring, productivity analysis, defect detection, and automated documentation [8,11,12,14]. Computer vision technologies have also great potential as field-based safety and health monitoring tools that can address limitations of current manual observational approaches, creating opportunities to automate the risk identification and evaluation processes by extracting and analyzing relevant information from images or videos [10,14,15]. Despite recent progress made toward computer vision-based approaches to safety and health monitoring, prior works have been infrequently applied to actual practice for technical and practical issues, or any other issues. Further studies are thus required to find out existing limitations and issues in the current body of knowledge in order to boost the adoption of the advanced techniques by practitioners as well as to address the identified issues in the future studies.

This paper reviews existing literature in computer vision-based safety and health monitoring: (1) to understand the existing state-of-the-art methods and their current progress; (2) to identify the major challenges and limitations commonly found in the prior studies; and (3) to offer potential problem-solving directions for future studies. To provide an overview on this review, this paper first presents current safety management practices relevant to field monitoring, and discusses potential roles and the general framework of computer vision for safety observation and inspections. We then present overall concepts, specific methods and applications of computer vision techniques that are directly or potentially used for safety and health monitoring in construction, specifically in the three groups categorized by computer vision methods and types of information extracted from imagery data (namely, object detection, object tracking and action recognition). Along with the overall review, this paper discusses technical challenges and potential issues when applying computer vision-based approaches in practice.

2. Overview of computer vision for safety and health monitoring

Compared with other industries such as automotive or manufacturing industries [62,63], the construction industry has many domain-specific issues in safety and health monitoring such as continuously changing and complex working environments, and non-standardized work procedures and designs, which may result in challenges for computer vision-based safety and health monitoring. In this section, current practices of safety and health monitoring in construction, and potential roles and approaches of computer vision to improve the practices are presented. Based on the potential roles, a general framework for computer vision-based safety and health monitoring is suggested.

2.1. Current practices of safety and health monitoring

The unique, dynamic, and complex nature of construction projects likely increases workers' exposure to hazardous working environments. Occupational hazards cannot be fully eliminated without systematic and comprehensive efforts for managing safety and health on construction worksites such as safety planning, worksite analysis, hazard prevention, and control or safety and health training [4]. The purpose of safety and health monitoring is to make sure that safety and health are being effectively managed by measuring health and safety practice against an organization's safety and health plans and standards.

Among safety and health monitoring activities, job safety observations and inspections are one of the common techniques used to evaluate ongoing tasks in construction [4]. The jobsite observation and inspection is typically conducted on a weekly or bi-weekly basis depending on the size of the project, and inspectors generally take the observation—without any other tasks at the time—for one to two hours at a randomly scheduled time during the week [16]. During the observation, the human observer then serves to detect and eliminate the potential causes (i.e., unsafe conditions and acts) of accidents by watching workers perform a specific task (i.e., safety observation) or visually examining the work area and work equipment (i.e., safety inspection) with a checklist [4]. To simplify the record keeping, for example, Reese and Eidson [4] identified unsafe conditions and acts that contribute to injury, property damage, or equipment failure, such as failure to wear personal protective equipment (PPE), improper lifting, improper use of equipment (e.g., excessive speeds, servicing moving equipment) or improperly stored explosive or hazardous materials.

2.2. Potential roles of computer vision-based approaches

As described in an earlier section, identification of risks such as unsafe conditions and acts calls upon observers' perceptual and cognitive capabilities [94]. For example, observers have to understand scenes using their perceptual capability such as object and scene recognition, or visual processing of spatial and temporal relations. Then, the perceptual information should be evaluated by comparing it with rules, guidelines, or observers' past experiences to identify unsafe conditions and acts. However, computer vision techniques themselves that aims to perform visual tasks accurately and reliably [17] are limited only to extract the perceptual information, not addressing evaluation of the information to identify unsafe conditions and acts. Therefore, computer vision-based approaches for safety and health monitoring should consider not only how to extract perceptual information, but also how to compare the information with existing knowledge on potential risks or hazards.

Types of perceptual information for safety and health monitoring vary depending on characteristics of unsafe conditions and acts, which requires different computer vision techniques. Based on the type of information, this paper classifies computer vision-based approaches into three categories: (1) scene-based; (2) location-based; and (3) action-based risk identification (Table 1). First, the scene-based approach pertains to understanding and evaluating any potential risk in a static scene by inspecting the scene in a safety context. For instance, required information for analysis includes whether unsafe objects are present in the scene, safety tools and equipment required are not present, workers are in unsafe areas, and so on. In Table 1, failure to wear PPE (e.g., safety vests and hard hats), congested work area, or improperly stored explosive or

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001

Table 1
Potential roles of computer vision techniques for identifying unsafe conditions and acts.

Approaches	Descriptions	Examples of targeted unsafe acts and conditions	Computer vision techniques applied
Scene-based risk identification	Identify unsafe acts and conditions by understanding static scenes at construction sites	Failure to wear personal protective equipment (PPE) Congested work areas Improperly stored explosive or hazardous materials, etc.	Object detection
Location-based risk identification	Identify unsafe acts based on locations and movements of project entities (e.g., equipment, workers)	Failure to warn coworkers against being struck by vehicles or equipment Improper working position Operating or working at unsafe speed etc.	Object tracking
Action-based risk identification	Identify violation of safety and health rules regarding motions	Improper movements of superstructure of heavy equipment Improper lifting with awkward postures, etc.	Action recognition

hazardous materials are examples of such unsafe acts and conditions. For these types of unsafe conditions and acts, object detection techniques have been applied in construction for the extraction of relevant information [18,43,46,47,49,61]. Object detection herein refers to the technique that detects a semantic object of interest in an image by searching the image with the known object model (e.g., appearance features). Therefore, object detection methods allow for assessing absence of PPE on workers by detecting the PPE and workers, or identifying unsafe conditions similarly by detecting hazardous materials at undesignated or known unsafe areas [14,18]. Second, the location-based approach is taken to evaluate the risk in the scene based on geometry information of project entities (e.g., equipment and workers) moving around over time. This type of information is essential to identify unsafe conditions and acts such as improper working position (e.g., close proximity between equipment or equipment and workers) or improper use of equipment (e.g., excessive speed of equipment). To gather the location information from images, previous research efforts in construction have applied object tracking techniques for tracking the trajectories of an object in a sequence of images by detecting the object on frame-by-frame basis [10,19–21]. At construction sites, construction workers and equipment continuously move over time to perform tasks, which may inadvertently cause accidents or injuries when workers or equipment operators fail to recognize the other workers or equipment present nearby. The location-based approach thus enables identifying interactions where the trajectories of workers and equipment frequently overlap; by anticipating potential sites of accidents, the location-bases approach allows managers to plan tasks properly (e.g., alarming dangerous situations and assigning a flagman or arranging the tasks at different time). Lastly, action-based risk identification focuses on the detection of unsafe action by equipment or workers at specific positions. For example, an improperly revolving superstructure of the excavator in congested areas is a common cause of struck-by-accidents [22]. In addition, unsafe actions by workers (e.g., improper lifting with awkward postures in Table 1) can cause ergonomic injuries (e.g., musculoskeletal disorders) which account for 24% of non-fatal occupational injuries and illnesses in construction in 2011 [2]. Motion information thus plays a critical role in evaluating ongoing works performed by equipment and workers, generally with existing safety and health rules (e.g., checklists). As a means to collect motion information on a jobsite, action recognition techniques that capture particular actions in sequential images have been studied in construction [24–29].

2.3. General framework for computer vision-based safety and health monitoring

Based on diverse approaches discussed above, we derived a general framework for computer vision-based safety and health monitoring as shown in Fig. 1. Computer vision-based approaches start from acquisition of imagery data on scene using various image sensing devices. To extract useful information for safety and health monitoring, the imagery data should be computationally analyzed using diverse computer vision techniques, which can be categorized into three elements according to types of information extracted from collected images or videos: (1) object detection; (2) object tracking; and (3) action recognition, as discussed in an earlier section.

Object detection serves as a method for scene-based risk identification, as well as a preliminary step for object tracking, and action recognition. Once project entities of interest are recognized

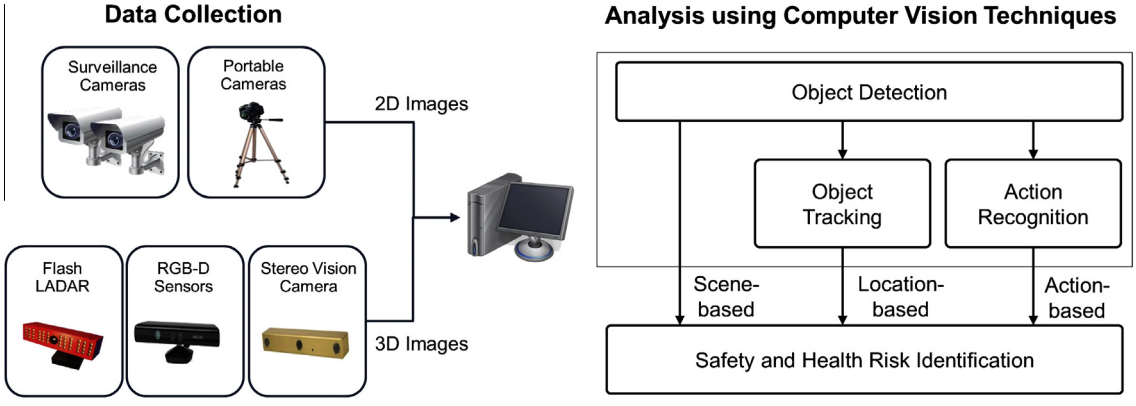

Fig. 1. General framework for computer vision-based safety and health monitoring.

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001

on a scene, their locations in 2D or 3D spaces can be tracked using object tracking algorithms from sequential images in the course of time. The location information extracted is used for location-based identification of unsafe conditions and acts. When the project entities (e.g., workers and equipment) with articulated body structures perform construction tasks, action recognition techniques are required to determine what a worker or equipment is doing from static or sequential images, and to identify unsafe acts that violate safety and health rules. Each category of computer vision techniques described here is a broad topic in the computer vision community, and many applications of the techniques have been studied in construction. A detailed review of current advances in each category will be described in the next section.

3. Computer vision-based approaches for construction safety and health monitoring

Based on literature review, this section provides detailed reviews on concepts, specific techniques, and applications of computer vision techniques in construction. Despite the current progress of computer vision studies in construction, studies in safety and health monitoring are still premature to be applied in real practice. While a large portion of studies have focused on data collection that can be used for multiple purposes such as performance monitoring including productivity analysis as well as safety and health monitoring, they fail to fully specify how to evaluate the data for detecting unsafe conditions and acts. However, as accurate and reliable data collection is also a necessary condition for safety and health monitoring, this paper reviews previous research efforts both on computer vision techniques that can be both directly and potentially used to collect required field data from images, and their applications on safety and health monitoring. In this section, first, we introduce diverse image sensing devices that are currently available for safety and health monitoring. Then, technical aspects, research challenges, and future direction of diverse computer vision approaches (i.e., object detection, object tracking and action recognition) are described. Table 2 presents an overview of previous research efforts to be reviewed in this section.

3.1. Imagery data collection

Computer vision-based safety and health monitoring requires images or videos on scenes where the construction task to be monitored is taking place. In construction, video cameras have been widely applied to observe and document construction activities on construction sites, and to provide project managers a live image of remote projects [30]. In addition, video surveillance systems using stationary or pan-tile-zoom cameras are common in construction for security issues such as accidents, vandalism, and the theft of materials [31]. These cameras can be inexpensive solutions to obtain 2D imagery data (i.e., 2D videos or sequential images) required for computer vision-based monitoring.

Recently, due to increased availability of new 3D data sensing devices, the use of 3D data becomes popular in computer vision applications [32]. Browatzki et al. [33] compared the performance of object detection based on 2D and 3D data respectively, as well as combination of 2D and 3D data, and found that 3D-based and combined approaches showed higher classification accuracy than the 2D data-based approach. The 3D data-based approach performs better because 3D data is less sensitive to lighting or color variances, contains geometrical cues, and provides better separation from the background [34]. In construction, laser-based 3D imaging systems and stereo-vision systems are generally used to collect real-time 3D data on construction sites [35,36,89,90]. For construction tasks conducted in indoor environments such as finishing work, RGB-D sensors such as MS Kinect™ can provide depth images that are generally used to analyze construction workers' postures and motions [27,28].

Table 3 shows comparison of diverse image sensing devices that can provide real-time data. Selection of image sensing devices should depend on the purpose of computer vision techniques applied. Chi and Caldas [21] suggested criteria on selection of appropriate image data collection devices for object identification and tracking on construction sites such as frame rate, outdoor application capability, reliable reading range, object localization capability and 3D modeling capability. Recently, the use of multiple cameras and 3D reconstruction techniques enables collecting 3D localization and modeling from 2D images [28,37,38]. However, as opposed to 2D cameras whose range varies depending on lens disparity and adjusted exposure values, selection of 3D imaging devices should be made more carefully because sensing accuracy is affected by distance from a camera and a scene [21,39], or lighting condition [36]. For example, Flash LADAR (Laser Detection and Ranging) and RGB-D sensors are usually unable to collect range data accurately in the presence of sunlight and have relatively short maximum range, which make it difficult to utilize these sensors on outdoor applications [28,40]. In addition, mobility and portability of devices need to also be considered because of continuously changing working spaces according to work progress. For example, even though stationary or pan-tilt-zoom cameras are already installed at construction sites, the cameras could not cover all scenes where construction tasks take place, and thus a potable camera with a tripod could be required to collect 2D images.

3.2. Object detection

To build physical structures, construction tasks involve numerous activities performed by multiple workers and equipment using various techniques and materials on a jobsite. To make computers understand a complex scene on construction sites, it is necessary to identify what kinds of project entities such as workers, equipment, and materials on the scene are involved for the tasks from the collected imagery data. Object detection can be directly used to identify unsafe conditions and acts at construction sites (e.g., failure to wear personal protective equipment and damaged area of building or defected material). For example, detection of safety gears such as safety vests and hard hats helps to distinguish construction workers from others (e.g., supervisors, engineers or pedestrians) [10,41–43,85,86], as well as to detect construction workers who do not wear safety gears for safety observation [18,87]. In addition, detection of construction workers and material surrounding equipment helps operators who have limited visibility on job sites to prevent collisions [14]. In addition, object detection in images or video streams acts as a first step for other computer vision-based approaches such as object tracking, and action recognition. For example, studies in construction have applied object detection algorithms to detect project entities such as construction workers, diverse construction equipment or multiple entities for initialization of objects to be tracked in a sequence of images [10,46,47] or for tracking objects frame-by-frame in sequential images [21,40,48].

3.2.1. Technical aspects of object detection

The most common approach for object detection is to divide the image window into small spatial regions, to extract features from the local windows, and then to classify the object of interest through supervised learning [92]. For an exhaustive search at different resolutions and locations within the image, a sliding window approach, which aims to find a bounding box around the object, is generally applied [91,93]. Once a potential sub-window is determined, image features, which are simplified representations

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001

Table 2
Overview of literature review on studies for computer vision-based safety and health monitoring.

Techniques	Types of data	Objective	Target objects
Object detection	2D images	Automated monitoring of work zone productivity or safety [61]	Construction equipment and worker
		Detection of construction workers not wearing a hard hat for safety [18]	Hard hat
		Detection of construction workers for initialization [43]	Multiple construction workers
		Automated vision-based monitoring system to detect off-highway dump trucks [46]	Multiple dump trucks
		Part-based object recognition model to detect a single excavator [47]	Single hydraulic excavator
		Kinematic key nodes model for dynamically detecting and locating movable objects [49]	Hydraulic excavator
Object tracking	2D images	Tracking construction workers for automated project performance control monitoring [19]	Construction worker
		Comparison of 2D vision tracking methods [20]	Construction equipment and worker materials
		Evaluation of the accuracy of vision and radio frequency tracking methods [79]	Multiple workers
		Tracking 3D locations of construction entities [38]	Construction equipment and worker
	3D images	Real-time 3D modeling using Flash LADAR to detect and track objects [39]	Static and dynamic objects
Object detection and tracking	2D images	Acquisition of 3D spatial coordinates of project entities [10]	Construction equipment and worker materials
		Detection and tracking of multiple workers from static and moving cameras [31]	Multiple workers
		Detection of construction workers and equipment while idle and in close proximity [48]	Construction equipment and worker
	2D images and range data	Automated image-based safety assessment method for earthmoving and surface mining activities [21]	Construction equipment
	3D images	Identification and tracking of objects surrounding the equipment [14]	Construction worker materials
Action recognition	2D images	Measurement of idle time of hydraulic excavators through action recognition [26]	Construction equipment
		Classification of actions by construction equipment and workers [24]	Construction equipment and worker
		Single action recognition of earthmoving equipment [25]	Construction equipment
		Vision-based monitoring framework for behavior-based safety management [29]	Construction worker
	3D images	Monitoring ergonomic risk on posture of construction workers [27]	Construction worker
	3D images and motion data	Unsafe action detection during ladder climbing [28]	Construction worker

Table 3
Comparison of image sensing devices.

Devices	Types of data	Available environment	Sensing range	Portability
Stationary camera	2D images	Both indoor and outdoor	Long	No
Potable camera	2D images	Both indoor and outdoor	Long	Yes
Flash LADAR	3D images	Indoor	Short (less than 10 m)	Yes
Stereo vision camera	2D + 3D images	Both indoor and outdoor	Long (accurate less than 10 m)	Yes
RGB-D sensor	3D images	Indoor	Short (less than 5 m)	Yes

of sub-window images, need to be extracted by selecting appropriate feature descriptors [32]. To extract features for 2D objects, dense Histogram of Oriented Gradients (HOG) [51], Local Binary Pattern (LBP) [52], Scale Invariant Feature Transform (SIFT) [54] have been proven successful for object classification [53]. One of the most popular methods for feature extraction for 3D data (i.e., point clouds) is the spin-image, which is used for surface representation [55]. Shape and geometry features [56] or Kernel descriptors [57] have been also used for depth images (i.e., RBG-D images) where spin image features do not always perform well [57]. Given the features, objects on 2D or 3D images can be classified using various methods. For example, a linear support vector machine approach has been widely used to provide an effective solution for classification problem, especially for histogram-based image classification [58]. To learn extensive datasets that require a large learning capacity, Convolutional Neural Networks (CNNs) have been proven to provide robust and effective classifiers [59]. Fig. 2 shows a conceptual illustration of overall procedures for object detection, which has been proven by Papageorgiou et al. [96].

For object detection, numerous algorithms have been developed as mentioned above, and results reveal that their performance highly relies on applications (e.g., objects) and conditions (e.g., light and viewpoint) where the algorithms are tested. Thus, previous studies in construction have tested existing algorithms under construction-specific conditions (e.g., moving objects), and try to identify optimal algorithms, which perform well for object detection in construction environments. As targets for object detection are generally moving in construction, background subtraction methods are widely used to find the rectangular foreground regions where the target objects to be classified exist, such that the exhaustive search which requires high computational effort is not required [41,43,46]. Because existing algorithms (e.g., a Mixtures of Gaussian method, a Codebook-based method, a Bayesian Model-based, and a median filter method) for moving object segmentation in 2D images shows different accuracy and computation time [41], previous studies selected algorithms that were suitable for their research purposes [43,46]. For example, Rezazadeh Azar and McCabe [46] applied the Bayesian Model for more accurate segmentation results while Park and Brilakis [43] focused on

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001

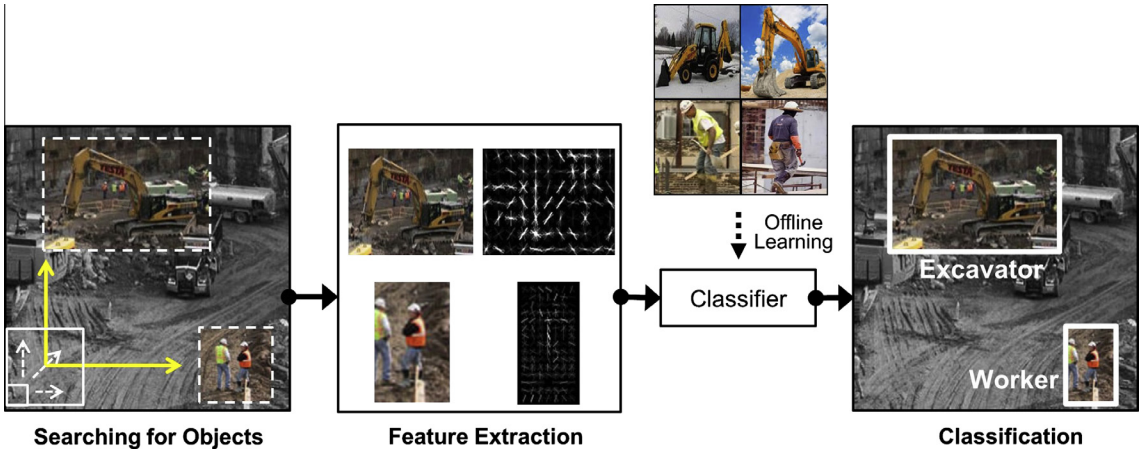

Fig. 2. Overall procedure for object detection.

computational efficiency, and thus selected the median filter method. To build a suitable 3D local model (i.e., target object) from range data, clustering algorithms such as a grid-based clustering algorithm [60] have been used [14]. For feature extraction, various types of features such as shape, color and motion features have been generally used, but there is a challenging issue to extract features that are compatible to various appearances of objects in construction [43]. As a shape-based feature, Haar-like and HOG features are widely applied for construction worker or equipment detection [43,47]. Chi and Caldas [61] combined shape-based features such as aspect ratio, height-normalized area size and percentage of occupancy of the bounding box, and a color-based feature such as average gray-scaled color of the area (bounding box) because they were not affected by the size of images. As a final step for object detection of project entities, previous studies applied various classifiers such as Support Vector Machine (SVM) [47], k-Nearest Neighbors (k-NN) [43] and Adaptive Boosting (AdaBoost) learning algorithm [46] to learn classifiers based on extracted features and to recognize project entities (e.g., equipment, workers) on images.

3.2.2. Research challenges and future study areas in object detection

One of the challenging issues for detection of project entities is classifying different shapes of project entities depending on viewpoints [47]. To address the issue regarding different visual orientations, Rezazadeh Azar and McCabe [47] attempted to learn a classifier using images with eight viewpoints of equipment, which have been showed robust performance in vehicle detection [64]. Another issue caused by equipment with kinematic joints is changing poses of equipment according to its movement. Yuan and Cai [49] found that the use of kinematic features can serve as deformable and semantic object features that represent the equipment's movement because construction equipment with kinematic joints has specific kinematic constraints from mechanic specification, and proposed the kinematic key nodes model for deformable object feature extraction and template training.

With respect to the object detection approach, previous studies in construction have made significant progress in the detection of specific types of objects (e.g., workers, equipment or damaged area) on scene images to address certain types of unsafe conditions (e.g., PPE). For comprehensive risk assessment, further research necessitates more attention on semantic understanding of the scene such as contextual relationships among objects within the scene. For example, to identify unsafe conditions associated with congested work areas at construction sites, several subtasks should be performed simultaneously such as detection of diverse project entities (e.g., workers, equipment and material), labeling of meaningful regions (e.g., designated area for stocking material, designated paths for workers and equipment, building structure), and even 3D reconstruction of the scene to obtain spatial information required to understand relationships between project entities and environments. These complex tasks are called "holistic scene understanding [65]" or "total scene understanding [66]." While detecting and classifying isolated objects and object classes is a critical component of object recognition, further studies based on these approaches are needed to be done to reach a complete understanding of visual scenes for safety and health monitoring at construction sites.

3.3. Object tracking

One of the key factors to determine unsafe conditions or acts at construction sites is the location of project entities as described in Section 2. Therefore, previous research efforts in construction have emphasized the need for automated tracking of project entities for real-time safety monitoring at construction sites. Compared with sensor-based tracking approaches such as Radio-Frequency Identification (RFID) tags, Ultra WideBand (UWB), and Global Positioning Systems (GPS), computer vision-based tracking methods have several advantages that: (1) it does not need to attach sensors on project entities and install signal receivers; (2) camera views can cover a large size of construction areas; and (3) multiple project entities on a scene can be tracked simultaneously [10]. Computer vision-based tracking approaches in construction have primarily focused on 2D-based tracking of project entities from video streams [19,20]. However, 2D locations of objects on images may not be enough to extract substantial information for the movement of project entities because proximity between objects or movements of objects (i.e., speed) cannot accurately be measured on a 2D plane, and thus it is difficult to determine whether the activities of project entities are safe or not without acquisition of 3D positions of project entities [10,38]. Therefore, 3D object tracking is essential in construction applications for safety and health monitoring of project entities.

3.3.1. Technical aspects of object tracking

Computer vision-based object tracking can create the temporal trajectory of detected objects as they move on a scene [50]. Once a vision tracker is initialized at the first frame of video streams through object detection algorithms, the tracking algorithm tracks the 2D projections of objects by assigning consistent labels to the tracked objects in a sequence of images from video streams. A

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001

Fig. 3. Overall procedure for object tracking.

conceptual illustration for object tracking, which have been proven by Yilmaz et al. [50] is shown in Fig. 3.

For 2D vision-based tracking, three types of methods have been widely used: (1) point tracking, (2) kernel tracking, and (3) silhouette tracking [50]. Point tracking uses feature points representing objects, and detect the object by matching the points in every frame. Kernel tracking tracks an object by computing the motion of the kernel that represents the object shape and appearance in consecutive frames. When the objects have complex shapes such as human, silhouette-based methods can provide more reliable tracking, compared with other tracking methods. The silhouette of the object is represented by a color histogram, object edges or contour, and thus silhouette trackers detect the object in each frame by shape matching and contour tracking. Studies related to 2D-based tracking in construction have focused on selection of appropriate algorithms and methods for construction application by comparing tracking performances for construction project entities such as workers and equipment. For example, Park et al. [20] performed a comparative study of 2D vision trackers (point-, kernel- and contour-based tracking) by evaluating their performance to determine the most effective solution for construction applications, and found kernel-based methods the most stable and reliable methods for tracking project entities in construction environments. Teizer and Vela [19] focused on the tracking of workers, and tested one kernel-based tracking method, mean-shift tracking [67], and three segmentation-based tracking methods such as Bayesian segmentation [68], active contours [69], and graph-cuts [70] that have been known to perform well for larger, non-rigid or variable targets. The algorithms were tested using diverse construction scenarios to assess tracking performance of four tracking methods and to identify their potential benefits and disadvantages. The overall results indicated that the Bayesian method performed best. Teizer and Vela [19] also found that contour-based methods may not be suitable for tracking a worker with variations in shapes depending on postures and camera positions, and the graph-cut method suffered from false positive image data.

For 3D object tracking, two types of approaches have been applied in construction applications: (1) triangulation-based 3D position tracking using multiple cameras [10,21,38], and (2) object tracking using 3D range sensors [39]. The triangulation-based method uses video streams from at least two cameras with different viewpoints [10,38]. This method basically starts from 2D object tracking of objects from two video streams. The tracked 2D trajectories are then converted into 3D coordinates by triangulation based on the fundamental matrix, focal length, and epipolar geometry, which associates two camera views. Despite the advantage of this approach, use of this method in real-time monitoring is limited due to computational time required to process the data [38]. Stereo vision camera systems (e.g., Bumblebee XB3, Point Grey Research, Inc.) that have laterally displaced multiple lenses and capability to compute depth information using triangulation in real time can be an alternative solution to reduce computing time [21]. Compared with the triangulation-based method, 3D range data has an advantage that 3D localization is straightforward once the object is detected. Among various types of 3D range sensing devices, flash LADAR [39] is suitable devices for real-time monitoring because they are portable and inexpensive devices with a fast frame rate (over 10 Hz). The 3D coordinates of project entities can provide information on the movement (e.g., path, speed, acceleration and direction) and location (e.g., proximity to other entities) necessary to identify unsafe conditions or acts at construction sites [10]. For example, the speed of equipment can be monitored to identify speed limit violations, and the path and location of equipment or workers can be used to detect dangerous access violation and close proximity violations between equipment or equipment and workers [21].

3.3.2. Research challenges and future study areas in object tracking

Considering that construction tasks frequently involve interactions between workers and equipment, safety and health monitoring requires simultaneous multiple targets tracking, which is challenging due to interacting trajectories causing occlusions and similar appearance of objects [31]. To address this issue, Yang et al. [31] suggested a robust multiple workforces tracking scheme using either stationary or pan-tilt-zoom cameras. Interactions between workers may result in tracking failure, which indicates that the workers are labeled wrongly. To manage the situations where workers are overlapped on images, Yang et al. [31] classified occlusion scenarios into three possible sequences: (1) no occlusion; (2) partial occlusion; and (3) severe occlusion. For the partial occlusion case, stepwise tracking (i.e., track the worker close to camera first, remove the pixels belonging the first target, and then track the other worker) was applied while the trajectories of workers were predicted by using the Kalman filter prediction when the workers are severely overlapped, and thus it is impossible to segment two workers on images. Although this approach showed robust tracking results on artificial scenarios, changing appearance of workers or severe occlusions on too many consequence frames are still challenging issues for multiple object tracking of workers.

The primary goal of studies on object tracking is to improve the accuracy of location estimation of project entities. In addition to the effectiveness of tracking algorithms, the accuracy can typically be affected by the distances between the cameras and project entities, and thus selection of the optimized camera location need to be considered to improve monitoring quality [21]. Also, camera calibration processes are of importance for large-scale projects as objects at various distances are seen and captured on a single image [71]. Furthermore, once accurate 3D locations of project entities are obtained, methods for displaying geometrical information to the interested users effectively should be investigated in

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001

future studies [10]. Brilakis et al. [10] suggested two possible solutions; display the movement of project entities in a virtual 3D space environment such as 3D CAD model, or superimpose trajectories on video images.

3.4. Action recognition

Object detection and location tracking provide useful data to observe and evaluate ongoing works and working environments based on the scene context (e.g., absence of personnel protective equipment and safety devices) and positional information (e.g., close proximity between workers and equipment). In certain situations, however, such information may not sufficiently allow for the comprehensive understanding of the scene in a safety context. For example, the proximity between an excavator and workers may not directly imply an unsafe condition if the excavator is not operating. On the other hand, when the excavator moves and rotates its boom and bucket to dig soil, workers working near the excavator are exposed to significant hazards if the operator of the excavator fails to recognize the workers. Furthermore, it is especially difficult to determine the risk of unsafe acts (e.g., reaching too far to one side on a ladder) that take place during a construction activity without motion and postural information; for instance, Occupational Safety and Health Administration (OSHA) recommends avoiding awkward postures (e.g., bending at the waist, twisting while lifting) to prevent ergonomic risk during material handling. These examples demonstrate the need for different types of information that help fully understand the scene (e.g., what a worker or equipment is doing) to identify particular unsafe acts and conditions. To this end, action recognition techniques have been applied to construction as a means to extract motion information from imagery data.

3.4.1. Technical aspects of action recognition

In the computer vision domain, human action recognition has been studied for the last few decades for many applications, including visual surveillance, video retrieval, and human–computer interaction. To analyze existing methods in detail, Poppe [72] conducted an extensive review on human action recognition in computer vision community. According to Poppe [72], human action recognition is the process of labeling images with action labels. It requires two steps: (1) image representation, and (2) action classification. Image representation for human action recognition is the extraction of human features (e.g., shapes and temporal motions) from images, which is conceptually similar to feature extraction for object detection. However, the features for action recognition should contain rich information, which is enough to classify diverse actions. Types of features for image representation can be divided into: (1) global representation which encodes the region of human body as a whole; (2) local representation which uses a collection of independent local patches; and (3) application-specific representation such as joint locations or joint angles from human skeletons, and accelerations of motions. Once images containing human action are represented, image features are classified as specific actions using diverse classification methods such as direct classification (e.g., k-NN and SVM classifiers), temporal state-space models (e.g., Hidden Markov models (HMM), Conditional random fields (CRF)), and detection-based methods (e.g., bag-of-words coding). Although not directly applicable, the procedures and algorithms for human action recognition can be also effectively used for equipment action recognition [25]. The overall procedure for action recognition is illustrated in Fig. 4.

In construction, action recognition has also been studied for various applications such as productivity and safety analysis by tracking the movements of project resources. Previous research efforts in the construction community have tested and further developed diverse vision-based approaches to both image representation and action classification, which are suitable to track and recognize motions of construction equipment and workers [24–29]. One of the commonly used approaches for image representations of equipment and workers is a silhouette-based feature, which exemplifies a global representation approach [26,27]. Zou and Kim [26] segmented the silhouette of an excavator in Hue, Saturation, and Value (HSV) color space, and calculated the centroid (the center of mass) pixel location of the silhouette to determine whether the excavator is moving or not by comparing the distance between the centroid in two consecutive images with the threshold value. Ray and Teizer [27] converted the range data of the region of interest (worker) obtained from a Kinect™ into grayscale values of person, and extracted a row of feature vectors from the grayscale image. After extracting feature vectors representing workers, poses of worker such as standing, squatting, stooping, and crawling were classified using Linear Discriminant Analysis (LDA) [27].

However, silhouette-based approaches that rely on abstract information of image features may generally perform well only for simply distinguishable actions and produces unreliable results sensitive to viewpoint, noise, and occlusions [72]. Taking into account that images taken from construction sites commonly contain significant noise and occlusions and that the objects of interest are articulated equipment (e.g., backhoes) and human bodies, local representations which represent images as a collection of local descriptors or patches may be more appropriate. Gong et al. [24] applied space–time interest point detector [73] to detect interest points on images of equipment and workers, and the HOG and Histogram of Optical Flow (HoF) descriptors to describe the interest points. To find representative features for each action, thousands of features were clustered using K-means algorithms, creating a code book of features, and then using Bayesian learning methods, actions in video streams were classified. Golparvar-Fard et al. [25] applied similar approaches with Gong et al. [24] by combining space–time interest point detectors [74] and local descriptor such as HOG [75]. In this paper, SVM classifier was learned using a set of spatio-temporal patterns of descriptors, called a code word to classify actions of equipment such as digging, hauling, dumping and swinging.

The approaches using image-based features have shown promising results and provided a valuable insight for action recognition of construction equipment and workers. For complex actions, however, the use of joint locations and joint angles can enhance the performance of action recognition by utilizing rich representation of actions [72]. The performance of this approach relies heavily on data collection of articulated objects, which is an ongoing challenge. To address this issue, recent studies have focused on motion capture techniques to extract human skeletons from imagery data. One of the widely used methods is a RGB-D sensor-based approach. A RGB-D sensor captures RGB-D images on scenes, and 3D human skeletons are extracted by using joint detection algorithms [27] or shape-fitting algorithms [28,88]. However, because the RGB-D sensor has a short operation range (about 4–5 m), and is sensitive to lighting conditions, previous studies are limited to indoor environments [27,28]. Han and Lee [29] suggested a state-of-the-art vision-based motion capture approach using multiple cameras that functions under less constraints on operating conditions. This approach applied 2D pose estimation algorithms to extract 2D skeletons from multiple viewpoints of images, and 3D reconstruction of 2D skeletons to obtain 3D skeletons. Compared with the RGB-D sensor-based approaches, the accuracy of the camera-based approach was, as yet, relatively low, but sufficient to recognize particular unsafe actions in the experiments [29]. Starbuck et al. [36] tried to combine merits of an RGB-D sensor-based approach (accuracy) and a camera-based

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001

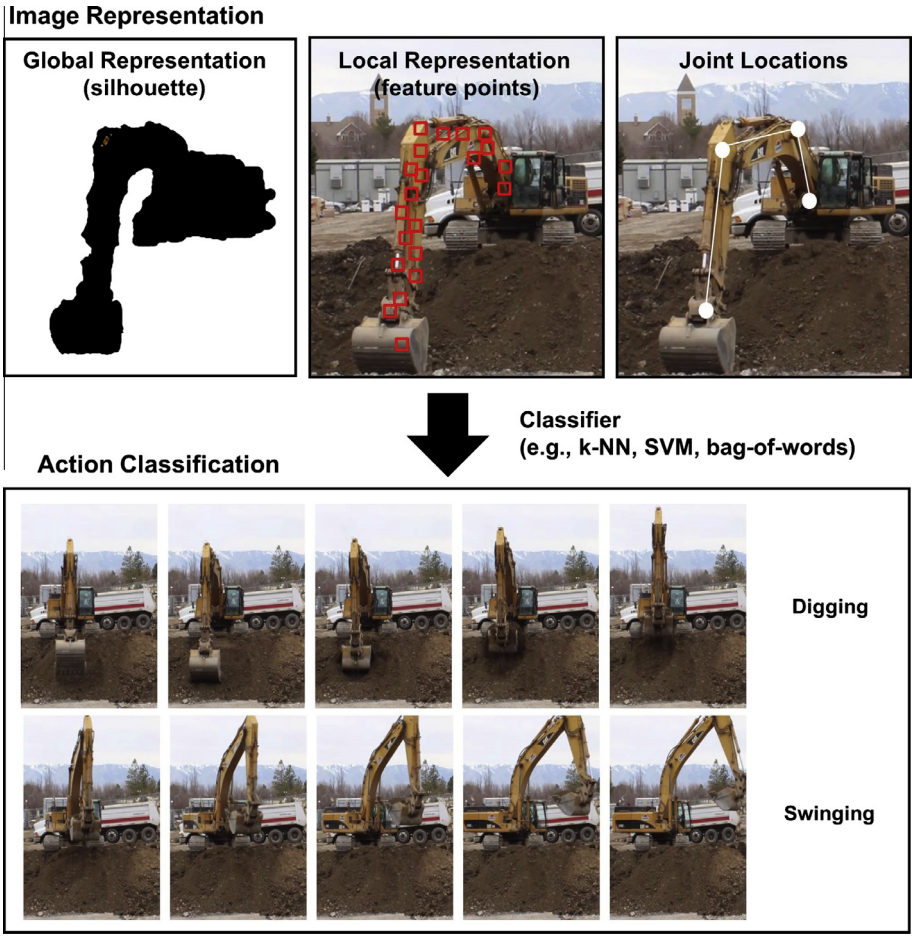

Fig. 4. Overall procedure for action recognition.

approach (long operation range) by suggesting a stereo vision camera-based motion capture approach. Unlike image-based action recognition methods that apply classification, human skeleton-based action recognition deals with actions as a series of movements of the human body, allowing more complicated action recognition [28,29].

Once actions by equipment and workers are recognized, the risk associated with the actions can be evaluated based on existing safety and health rules. With regard to equipment, the combination of location information from object tracking and operating status (idling and operating) from action recognition can provide comprehensive understanding on potential unsafe acts or conditions. Construction workers are exposed to both safety and health risks according to their actions, and therefore different criteria are required for safety and health monitoring respectively. Han et al. [28] and Han and Lee [29] defined unsafe actions from literature review, and by comparing defined unsafe action templates with workers' motion data over time, unsafe actions by workers are automatically detected. Ray and Teizer [27] evaluated ergonomic risks based on body angles during performing manual tasks by using posture-based ergonomic analysis methods. For more fundamental ergonomic risk analysis of workers' postures and movements, Seo et al. [76] proposed motion-data driven biomechanical analysis for identification of tasks generating excessive musculoskeletal stresses.

3.4.2. Research challenges and future study areas in action recognition

Action recognition is one of the more challenging topics in computer vision due to large variations and complexity in movements that change over time. Golparvar-Fard et al. [25] investigated current challenging issues on the state-of-the-art computer vision-based action recognition in construction applications—those are: (1) lack of datasets for action recognition of diverse types of equipment; (2) complex and articulated actions of construction equipment and workers; (3) lack of knowledge to define a time-series of actions (i.e., starting point and duration); (4) simultaneous action recognition of multiple project entities; and (5) lack of a holistic approach to benchmarking, monitoring, and visualization of performance information.

Several studies have attempted to address these issues. For example, with regard to datasets, Golparvar-Fard et al. [25] presented the first comprehensive video dataset for action recognition of excavators and dump trucks. Articulated actions by equipment and workers have been addressed in several studies by using local representation of image features [24,25] and skeleton-based motion data [28,29]. The use of dynamic time warping (DTW) can be a solution for recognizing a time-series of actions with different times and paces because of the capacity to measure a distance between action samples with different lengths to detect specific actions of interest among sequential actions [28]. However, several research problems such as action recognition for multiple equipment/workers and a holistic approach of action recognition for practitioners have not been fully solved yet, and thus further studies are required for achieving robust automatic action recognition to monitor unsafe actions in practice.

4. Discussion

In this section, we discuss technical and practical issues that commonly arise when applying diverse computer vision

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001

techniques for safety and health monitoring at real construction sites, and identify potential solutions and future directions to address the issues. These include a lack of task-specific and quantitative metrics for evaluation of unsafe conditions and acts, technical obstacles (e.g., occlusions, proper camera positions and datasets) due to dynamic conditions at construction sites, and privacy issues related to continuous monitoring.

4.1. Lack of task-specific and quantitative metrics to evaluate unsafe conditions and acts

Computer vision techniques used for safety and health monitoring have been mainly focused on extracting useful information such as presence/absence of objects of interest on scenes, location information, and action categories. Based on the information, evaluation of safety and health risks can be made by comparing current status with existing safety and health rules (e.g., checklist). However, because existing checklists contain general rules for construction safety and health and are assessed by subjective evaluation based on safety personnel's experience, those checklists cannot be directly used for computer vision-based monitoring for specific situations. For example, while unsafe acts such as failure to wear personal protective equipment (PPE) can be assessed only by detecting PPE on images, identification of some unsafe conditions and acts require specific rules for proper operation speed at diverse site conditions and acceptable proximity between objects (e.g., excessive speeds of equipment and improper working position).

Identifying specific and quantitative rules to evaluate construction tasks should therefore be preceded for computer vision-based safety and health monitoring, but only a few studies have addressed this issue. For example, Chi and Caldas [21] identified the possible causes of accidents during earthmoving and surface mining activities from literature, and suggested specific spatial needs for safety assessment of equipment operation (e.g., high operation speed, access to dangerous areas, and close proximity between objects) based on current safety regulations, academic safety studies, industrial standards and discussion with industrial experts. Similarly, Han and Lee [29] investigated a list of critical body postures and motions that may lead to traumatic or ergonomic injuries from the OSHA accident statistics, training materials, and checklists in the previous studies, and then evaluated workers' actions to determine whether the actions are unsafe or not. Specific ergonomic rules such as NIOSH lifting equation [77] were also used to evaluate workers' posture quantitatively during lifting tasks [27]. These studies imply that there is no commonly acceptable rule available for computer vision-based risk identification during construction tasks, and specific and quantifiable rules should be pre-identified for target tasks to be evaluated from diverse sources. Thus, for practical application of computer vision-based safety and health monitoring for diverse construction tasks, further studies should be completed to identify task-specific and quantitative metrics for construction task evaluation.

4.2. Obstacles due to dynamic conditions at construction sites

The pre-requisite for successful safety and health monitoring is the reliability and accuracy of data collected by computer vision techniques, which is challenging due to unique characteristics of construction. Construction is characterized by its dynamics such as worksites involving multiple workers, diverse types of equipment and material, and continuously changing working environments. These dynamic features at construction sites cause several technical issues for computer vision application, such as occlusion, the selection of appropriate camera positions, and the need for comprehensive image datasets with diverse viewpoints.

4.2.1. Occlusion

Occlusions are one of the major issues when applying computer vision techniques for construction environments. Particularly, construction activities involving numerous equipment and workers taking place often in a congested space inevitably generate a significant level of occlusions (i.e., self-occlusion, interobject occlusion, and occlusion by the background scene structure) [50]. To address this issue, Yang et al. [31] tested diverse occlusion scenarios when tracking project entities by suggesting occlusion management scheme. However, this paper found that most studies in construction ignored occlusions by assuming no occlusion in the scene. Resolving occlusions is a critical topic in computer vision community. Basically, if the object of interest is invisible on the scene, it is impossible to detect and track the object. However, if occlusions occur temporarily or partially, several algorithmic solutions exist. For example, for resolving partial occlusions due to inter-occluded humans, Wu and Nevatia [78] suggested a part-based representation for humans using edgelet features, and explicitly model the inter-object occlusion. This approach showed a robust detection result on a partially occluded human. Even when the object of interest is completely lost due to occlusions by the background scene structure, there are several ways to recover the missing trajectories. One of the common methods for complete occlusion for object tracking is to predict the trajectory of the object by linear dynamic or nonlinear dynamic models [50]. A Kalman filter that was used by Yang et al. [31] is an example of this approach for estimating the location and motion of the object. Combining vision-based approaches with sensor-based approaches can be also a practical solution to cope with occlusions as well as to enhance performance, considering that vision-based approaches are sensitive to the environments (e.g., lighting conditions and occlusions) where construction tasks are performed [25]. Yang et al. [79] compared a vision-based tracking and a commercial UWB tracking system in diverse construction scenarios, and found that the vision-based tracking became more reliable and accurate when combined with the UWB system, compared with the scenario where each approach was applied respectively.

4.2.2. Selection of appropriate camera positions

Although video or surveillance cameras are increasingly prevalent on construction sites, the use of these cameras for safety and health monitoring can be limited due to: (1) continuously changing working spaces for specific construction tasks, (2) limited accessibility to the working spaces, and (3) frequent occlusions when multiple tasks are conducted in a limited space. For these reasons, the number and position of cameras should be carefully determined prior to work. In practice, a preconstruction conference to assess potential hazards which may arise during the project is an essential step for safety and health management. Generally, subcontractors prepare a safety plan that contains possible serious hazards and their locations [4]. Based on the information, types of tasks to be monitored and locations where the tasks are performed should be determined prior to construction. However, numbers and positions of camera required for vision-based monitoring vary depending on computer vision algorithms applied because most of studies in construction assumed specific site conditions and camera positions. In other words, most previous studies reviewed in this paper tested and validated the proposed method in experiments rather than in an actual site. To make computer vision-based monitoring more practical, more studies need to be performed to test proposed methods in diverse field settings, and thus suggest a guideline for recording settings of cameras, considering various site conditions. Harsh conditions may arise where the installation of camera that covers the whole view of the scene of tasks is impossible (e.g., large earthmoving work), or the access to the site is limited (e.g., roofing, bridge construction). Recently,

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001

Unmanned Aerial Vehicles (UAVs) that operate under remote/autonomous control are considered as alternative devices to collect real-time video of the current situation on construction sites [80,95]. Further research efforts are therefore required to evaluate the feasibility of UAVs to collect imagery data for vision-based safety and health monitoring without spatial constraints.

4.2.3. Comprehensive image datasets with diverse viewpoints

There are no comprehensive datasets publicly available for computer vision studies in construction. Previous studies in construction have used their own datasets, which may result in biased algorithms to a particular dataset [72]. Considering the dynamics in construction where diverse types of equipment are involved, and various actions by workers and equipment exist, larger and more complex image datasets from various viewpoints should be required to guide research efforts to practical and realistic direction. The use of common datasets for computer vision studies in construction is also important for benchmarking and evaluating developed computer vision algorithms [25]. Golparvar-Fard et al. [25] emphasized the need for comprehensive datasets in construction domain, and introduced a new dataset for action recognition for commonly used earthmoving equipment (dump trucks and excavators). However, more research work should be done on the creation of publicly available datasets that are not only application-specific (e.g., object detection and action recognition), but also contain sufficient variation on types of equipment, viewpoints, and actions.

4.3. Privacy issue due to continuous monitoring at construction sites

The rapidly developing computing environment can now collect 'context-aware' information about what, when and where workers do [81]; this invokes controversy as to whether automated monitoring of workers using cameras or location tracking devices should be allowed or not [82]. Some support automated workplace surveillance due to organizational benefits, while others criticize it because it can increase stress in workers by essentially dehumanizing workers by invading worker privacy [82]. In particular, videos required for computer vision-based monitoring contain privacy-intrusive data, and thus vision-based monitoring at construction sites may cause undesirable effects to workers. Ball [83] insisted that excessive monitoring of workers at workplaces can be detrimental to workers for a number of reasons: (1) privacy concerns when workers' information is disclosed without their consent; (2) restriction of creative behaviors if workers realize their actions are monitored; and (3) creation of 'anticipatory conformity' where workers perform a task in a accepting way, reducing the amount of their commitment and motivation.

To mediate the negative effects of vision-based monitoring, careful managerial considerations are required. For example, because secret monitoring may create negative perceptions on monitoring procedures and result in the loss of trust in management, supervisors should not attempt to monitor a worker from a concealed position [84]. The worker should be aware that he or she will be monitored using video cameras. All employees must understand the reason for vision-based safety and health monitoring, and that it helps reduce injuries by eliminating the potential causes of accidents [4].

5. Conclusion

This paper presented an extensive review of computer vision-based approaches for safety and health monitoring at construction sites. Based on the current practice for safety and health monitoring, we derived potential roles of computer vision techniques for identifying unsafe acts and conditions, and classified them into: (1) scene-based; (2) location-based; and (3) action-based risk identification that can be achieved by specific techniques such as object detection, object tracking, and action recognition respectively. Object detection is not only a preliminary process for object tracking and action recognition, but also a method for scene-based identification of safety risks such as failure to wear PPE or damaged area of structure. Object tracking is the process to track 2D or 3D trajectories of project entities such as equipment and workers, and provides information on locations and movements of project entities that can be used to identify unsafe acts or conditions such as speed limit violations of equipment, close proximity between equipment or between equipment and workers. Action recognition provides rich information on postures and motions of equipment and workers, and thus unsafe acts can be automatically assessed.

This paper provided detailed summaries of technical aspects in each category, and their applications in the construction community along with technical challenges that previous studies have not been fully addressed yet. Despite the recent progress in computer vision-based approaches, lots of studies are still focusing on field data collection that is directly or potentially used for safety and health monitoring, having several technical challenges. The challenges include: (1) comprehensive risk assessment through total scene understanding; (2) improvement of tracking accuracy by selecting appropriate camera positions and effective visualization of tracking results; and (3) simultaneous recognition of actions by multiple equipment/workers and action recognition in a holistic manner for practitioners.

In addition, we discussed technical and practical issues that may arise when applying computer vision-based safety and health monitoring in practice. From the literature review, we need task-specific and quantitative metrics to evaluate unsafe conditions and acts based on information extracted from images because there is no commonly acceptable rule that can be used for computer vision-based risk identification. In addition, due to continuously changing working environments and diverse project entities involved in construction, several issues such as occlusion, selection of appropriate camera positions, and needs for comprehensive image datasets with diverse viewpoints should be addressed in future studies for practicality of computer vision-based approaches. Continuous monitoring can also cause negative effects such as privacy concerns, restriction of creative behavior, and decreased motivation by construction workers; thus practitioners should consider managerial (e.g., consent from workers and education) or technical solutions (e.g., privacy preserving techniques) to mitigate the detrimental results.

Considering the current technical advancement and potential benefits from computer vision-based approaches, it is believed that these challenges and issues will be solved in the near future. We also expect that this paper can provide not only valuable insight to researchers who work on computer vision-based safety and health monitoring, but also opportunities and considerations to practitioners who want to apply computer vision techniques in practice.

Acknowledgements

The work presented in this paper was supported financially with a National Science Foundation Award (No. CMMI-1161123). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Bureau of Labor Statistics, National Census of Fatal Occupational Injuries in 2012. USDL-13-1699, 2012a. <http://www.bls.gov/news.release/cfoi.nr0.htm>.

[2] Bureau of Labor Statistics, Nonfatal Occupational Injuries and Illnesses Requiring Days Away From Work, 2012. USDL-13-2257, 2012b. <http://www.bls.gov/news.release/osh2.nr0.htm>.
[3] K. Ringen, J. Seegal, A. England, Safety and health in the construction industry, Ann. Rev. Public Health 16 (1) (1995) 165–188.
[4] C.D. Reese, J.V. Eidson, Handbook of OSHA Construction Safety and Health, CRC Press, 2006.
[5] H.W. Heinrich, Industrial Accident Prevention, McGraw-Hill Book Company, New York, 1959.
[6] J. Hinze, R. Godfrey, An evaluation of safety performance measures for construction projects, J. Constr. Res. 4 (01) (2003) 5–15.
[7] R.E. Levitt, Construction Safety Management, John Wiley & Sons, 1993.
[8] S. Taneja, B. Akinci, J.H. Garrett, L. Soibelman, E. Ergen, A. Pradhan, E.B. Anil, Sensing and field data capture for construction and facility operations, J. Constr. Eng. Manage. 137 (10) (2011) 870–881.
[9] H. Laitinen, M. Marjamäki, K. Päivärinta, The validity of the TR safety observation method on building construction, Accid. Anal. Prevent. 31 (5) (1999) 463–472.
[10] I. Brilakis, M.W. Park, G. Jog, Automated vision tracking of project related entities, Adv. Eng. Inform. 25 (4) (2011) 713–724.
[11] M. Golparvar-Fard, F. Peña-Mora, Application of visualization techniques for construction progress monitoring, Comput. Civil Eng. 2007 (2007) 27.
[12] J. Gong, C.H. Caldas, Computer vision-based video interpretation model for automated productivity analysis of construction operations, J. Comput. Civil Eng. 24 (3) (2009) 252–263.
[13] S.M. Shahandashti, S.N. Razavi, L. Soibelman, M. Berges, C.H. Caldas, I. Brilakis, J. Teizer, P.A. Vela, C.T. Haas, J. Garrett, B. Akinci, Z. Zhu, Data fusion approaches and applications for construction engineering, ASCE J. Constr. Eng. Manage. Reston Virginia 137 (10) (2011) 863–869.
[14] S. Chi, C.H. Caldas, D.Y. Kim, A methodology for object identification and tracking in construction based on spatial modeling and image matching techniques, Comput.-Aid. Civil Inf. Eng. 24 (3) (2009) 199–211.
[15] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, F. Peña-Mora, Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques, Automat. Constr. 20 (8) (2011) 1143–1155.
[16] J. Hinze, Construction Safety, Prentice-Hall, Upper Saddle River, NJ, 1997.
[17] J.H. Elder, R.M. Goldberg, Image editing in the contour domain, IEEE Trans. Pattern Anal. Mach. Intell. 23 (3) (2001) 291–296.
[18] S. Du, M. Shehata, W. Badawy, Hard hat detection in video sequences based on face features, motion and color information, 2011 3rd International Conference on Computer Research and Development (ICCRD), vol. 4, IEEE, 2011, pp. 25–29.
[19] J. Teizer, P.A. Vela, Personnel tracking on construction sites using video cameras, Adv. Eng. Inform. 23 (4) (2009) 452–462.
[20] M.W. Park, A. Makhmalbaf, I. Brilakis, Comparative study of vision tracking methods for tracking of construction site resources, Automat. Constr. 20 (7) (2011) 905–915.
[21] S. Chi, C.H. Caldas, Image-based safety assessment: automated spatial safety risk identification of earthmoving and surface mining activities, J. Constr. Eng. Manage. 138 (3) (2012) 341–351.
[22] NIOSH, Preventing Injuries When Working with Hydraulic Excavators and Backhoe Loaders, Publication No. 2004-107, The National Institute for Occupational Health and Safety, Cincinnati, OH, 2003. <http://www.cdc.gov/niosh/docs/wp-solutions/2004-107/pdfs/2004-107.pdf> (accessed September 2014).
[23] N. Pradhananga, J. Teizer, Automatic spatio-temporal analysis of construction equipment operations using GPS data, Autom. Constr. 29 (2013) 107–122.
[24] J. Gong, C.H. Caldas, C. Gordon, Learning and classifying actions of construction workers and equipment using Bag-of-Video-Feature-Words and Bayesian network models, Adv. Eng. Inform. 25 (4) (2011) 771–782.
[25] M. Golparvar-Fard, A. Heydarian, J.C. Niebles, Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers, Adv. Eng. Inform. 27 (4) (2013) 652–663.
[26] J. Zou, H. Kim, Using hue, saturation, and value color space for hydraulic excavator idle time analysis, J. Comput. Civil Eng. 21 (4) (2007) 238–246.
[27] S.J. Ray, J. Teizer, Real-time construction worker posture analysis for ergonomics training, Adv. Eng. Inform. 26 (2) (2012) 439–455.
[28] S. Han, S. Lee, F. Peña-Mora, Vision-based detection of unsafe actions of a construction worker: case study of ladder climbing, J. Comput. Civil Eng. 27 (6) (2012) 635–644.
[29] S. Han, S. Lee, A vision-based motion capture and recognition framework for behavior-based safety management, Automat. Constr. 35 (2013) 131–141.
[30] J.G. Everett, H. Halkali, T.G. Schlaff, Time-lapse video applications for construction project management, J. Constr. Eng. Manage. 124 (3) (1998) 204–209.
[31] J. Yang, O. Arif, P.A. Vela, J. Teizer, Z. Shi, Tracking multiple workers on construction sites using video cameras, Adv. Eng. Inform. 24 (4) (2010) 428–434.
[32] M. Baum, Using spinImages for 3D Object Classification. Bachelor Thesis, Faculty of Technology, Bielefeld University, 2011.
[33] B. Browatzki, J. Fischer, B. Graf, H.H. Bulthoff, C. Wallraven, Going into depth: evaluating 2D and 3D cues for object classification on a new, large-scale object dataset, in: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), IEEE, 2011, pp. 1189–1195.
[34] R. Socher, B. Huval, B. Bath, C.D. Manning, A. Ng, Convolutional-recursive deep learning for 3D object classification, in: Advances in Neural Information Processing Systems, 2012, pp. 665–673.
[35] Y.K. Cho, C. Wang, P. Tang, C.T. Haas, Target-focused local workspace modeling for construction automation applications, J. Comput. Civil Eng. 26 (5) (2011) 661–670.
[36] R. Starbuck, J. Seo, S. Han, S. Lee, A stereo vision-based approach to marker-less motion capture for on-site kinematic modeling of construction worker tasks, in: Proceedings of the 15th International Conference on Computing in Civil and Building Engineering (ICCCBE), Orlando, FL, 2014.
[37] M. Golparvar-Fard, F. Peña-Mora, C.A. Arboleda, S. Lee, Visualization of construction progress monitoring with 4D simulation model overlaid on time-lapsed photographs, J. Comput. Civil Eng. 23 (6) (2009) 391–404.
[38] M.W. Park, C. Koch, I. Brilakis, Three-dimensional tracking of construction resources using an on-site camera system, J. Comput. Civil Eng. 26 (4) (2012) 541–549.
[39] J. Teizer, C.H. Caldas, C.T. Haas, Real-time three-dimensional occupancy grid modeling for the detection and tracking of construction resources, J. Constr. Eng. Manage. 133 (11) (2007) 880–888.
[40] D. Anderson, H. Herman, A. Kelly, Experimental characterization of commercial flash ladar devices, in: International Conference of Sensing and Technology, vol. 2, 2005.
[41] J. Gong, C.H. Caldas, An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations, Automat. Constr. 20 (8) (2011) 1211–1226.
[42] G. Gualdi, A. Prati, R. Cucchiara, Contextual information and covariance descriptors for people surveillance. an application for safety of construction workers, J. Image Video Process. 2011 (2011) 9.
[43] M.W. Park, I. Brilakis, Construction worker detection in video frames for initializing vision trackers, Automat. Constr. 28 (2012) 15–25.
[44] J. Teizer, B.S. Allread, C.E. Fullerton, J. Hinze, Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system, Automat. Constr. 19 (5) (2010) 630–640.
[45] T. Cheng, U. Mantripragada, J. Teizer, P.A. Vela, Automated trajectory and path planning analysis based on ultra wideband data, ASCE J. Comput. Civil Eng. Reston Virginia 26 (2012) 151–160.
[46] E. Rezazadeh Azar, B. McCabe, Automated visual recognition of dump trucks in construction videos, J. Comput. Civil Eng. 26 (6) (2012) 769–781.
[47] E. Rezazadeh Azar, B. McCabe, Part based model and spatial–temporal reasoning to recognize hydraulic excavators in construction images and videos, Automat. Constr. 24 (2012) 194–202.
[48] M. Memarzadeh, M. Golparvar-Fard, J.C. Niebles, Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors, Automat. Constr. 32 (2013) 24–37.
[49] C. Yuan, H. Cai, Key nodes modeling for object detection and location on construction site using color-depth cameras, in: Proceedings of the 15th International Conference on Computing in Civil and Building Engineering (ICCCBE), Orlando, FL, 2014.
[50] A. Yilmaz, O. Javed, M. Shah, Object tracking: a survey, Acm Comput. Surv. (CSUR) 38 (4) (2006) 13.
[51] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005, vol. 1, IEEE, 2005, pp. 886–893.
[52] X. Wang, T.X. Han, S. Yan, An HOG-LBP human detector with partial occlusion handling, in: 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 32–39.
[53] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, T. Huang, Large-scale image classification: fast feature extraction and svm training, in: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2011, pp. 1689–1696.
[54] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[55] A. Johnson, Spin-Images: A Representation for 3-D Surface Matching. PhD thesis, Robotics Institute, Carnegie Mellon University, 1997.
[56] H.S. Koppula, A. Anand, T. Joachims, A. Saxena, Semantic labeling of 3d point clouds for indoor scenes, in: Advances in Neural Information Processing Systems, 2011, pp. 244–252.
[57] L. Bo, X. Ren, D. Fox, Depth kernel descriptors for object recognition, in: 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2011, pp. 821–826.
[58] O. Chapelle, P. Haffner, V.N. Vapnik, Support vector machines for histogram-based image classification, IEEE Trans. Neural Networks 10 (5) (1999) 1055–1064.
[59] A. Krizhevsky, I. Sutskever, G. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 1106–1114.
[60] P. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Pearson Education Inc, Boston, MA, 2005.
[61] S. Chi, C.H. Caldas, Automated object identification using optical video cameras on construction sites, Comput. Aided Civ. Inf. Eng. 26 (5) (2011) 368–380.
[62] B. Coifman, D. Beymer, P. McLauchlan, J. Malik, A real-time computer vision system for vehicle tracking and traffic surveillance, Transport. Res. Part C: Emer. Technol. 6 (4) (1998) 271–288.
[63] T. Brosnan, D.W. Sun, Improving quality inspection of food products by computer vision––a review, J. Food Eng. 61 (1) (2004) 3–16.

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001

[64] P.E. Rybski, D. Huber, D.D. Morris, R. Hoffman, Visual classification of coarse vehicle orientation using histogram of oriented gradients features, in: IEEE Intelligent Vehicles Symposium, IEEE, New York, 2010, pp. 921–928.
[65] G. Heitz, S. Gould, A. Saxena, D. Koller, Cascaded classification models: combining models for holistic scene understanding, in: Advances in Neural Information Processing Systems, 2009, pp. 641–648.
[66] L.J. Li, R. Socher, L. Fei-Fei, Towards total scene understanding: classification, annotation and segmentation in an automatic framework, in: IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009, IEEE, 2009, pp. 2036–2043.
[67] D. Comaniciu, V. Ramesh, P. Meer, Kernel-based object tracking, IEEE Trans. Pattern Anal. Mach. Intell. 25 (5) (2003) 564–577.
[68] P.A. Vela, M. Niethammer, G.D. Pryor, A.R. Tannenbaum, R. Butts, D. Washburn, Knowledge-based segmentation for tracking through deep turbulence, IEEE Trans. Control Syst. Technol. 16 (3) (2008) 469–474.
[69] D. Freedman, T. Zhang, Active contours for tracking distributions, IEEE Trans. Image Process. 13 (4) (2004) 518–526.
[70] J. Malcolm, Y. Rathi, A. Tannenbaum, Multi-object tracking through clutter using graph cuts, in: IEEE 11th International Conference on Computer Vision, 2007, ICCV 2007, IEEE, 2007, pp. 1–5.
[71] H. Fathi, I. Brilakis, A transformational approach to explicit stereo camera calibration for improved euclidean accuracy of infrastructure 3D reconstruction, in: 2013 ASCE International Workshop on Computing in Civil Engineering, 2013.
[72] R. Poppe, A survey on vision-based human action recognition, Image Vision Comput. 28 (6) (2010) 976–990.
[73] I. Laptev, On space-time interest points, Int. J. Comput. Vis. 64 (2–3) (2005) 107–123.
[74] J.C. Niebles, H. Wang, L. Fei-Fei, Unsupervised learning of human action categories using spatial-temporal words, Int. J. Comput. Vis. 79 (3) (2008) 299–318.
[75] I. Laptev, M. Marszalek, C. Schmid, B. Rozenfeld, Learning realistic human actions from movies, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008, IEEE, 2008, pp. 1–8.
[76] J. Seo, R. Starbuck, S. Han, S. Lee, T.J. Armstrong, Motion data-driven biomechanical analysis during construction tasks on sites, J. Comput. Civil Eng. (2014), http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000400. B4014005.
[77] T.R. Waters, V. Putz-Anderson, A. Garg, L.J. Fine, Revised NIOSH equation for the design and evaluation of manual lifting tasks, Ergonomics 36 (7) (1993) 749–776.
[78] B. Wu, R. Nevatia, Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors, in: Tenth IEEE International Conference on Computer Vision, 2005, ICCV 2005, vol. 1, IEEE, 2005, pp. 90–97.
[79] J. Yang, T. Cheng, J. Teizer, P.A. Vela, Z.K. Shi, A performance evaluation of vision and radio frequency tracking methods for interacting workforce, Adv. Eng. Inform. 25 (4) (2011) 736–747.
[80] M. Gheisari, J. Irizarry, B.N. Walker, UAS4SAFETY: the potential of unmanned aerial systems for construction safety applications, in: Construction Research Congress 2014 Construction in a Global Network, ASCE, pp. 1801–1810.
[81] H. Bristow, C. Baber, J.F. Knight, S.I. Woolley, Defining and evaluating context for wearable computing, Int. J. Human–Comput. Stud. 60 (5–6) (2004) 798–819.
[82] G.S. Alder, Ethical issues in electronic performance monitoring: a consideration of deontological and teleological perspectives, J. Bus. Ethics 17 (7) (1998) 729–743.
[83] K. Ball, Workplace surveillance. An overview, Labor Hist. 51 (1) (2010) 87–106.
[84] F. Tabak, W.P. Smith, Privacy and electronic monitoring in the workplace. A model of managerial cognition and relational trust development, Emp. Respons. Rights J. 17 (3) (2005) 173–189.
[85] T.I.P. Weerasinghe, J.Y. Ruwanpura, Automated data acquisition system to assess construction worker performance, in: Proceedings of Construction Research Congress 2009, 2009, pp. 11–20.
[86] T.I.P. Weerasinghe, J.Y. Ruwanpura, Automated multiple objects tracking system (AMOTS), in: Construction Research Congress 2010, 2010, pp. 11–20.
[87] T.I.P. Weerasinghe, Automated Construction Worker Performance and Tool-time Measuring Model Using RGB Depth Camera and Audio Microphone Array System, PhD Thesis, University of Calgary, 2013. <http://hdl.handle.net/11023/605>.
[88] A. Khosrowpour, J.C. Niebles, M. Golparvar-Fard, Vision-based workface assessment using depth images for activity analysis of interior construction operations, Automat. Constr. 48 (2014) 74–87.
[89] C. Wang, Y.K. Cho, Smart scanning and near real-time 3D surface modeling of dynamic construction equipment from a point cloud, Automat. Constr. 49 (2014) 239–249.
[90] Y.K. Cho, M. Gai, Projection-recognition-projection method for automatic object recognition and registration for dynamic heavy equipment operations, J. Comput. Civil Eng. 28 (2014). SPECIAL ISSUE: 2012 International Conference on Computing in Civil Engineering, A4014002.
[91] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell. 32 (9) (2010) 1627–1645.
[92] K. Murphy, A. Torralba, D. Eaton, W. Freeman, Object detection and localization using local and global features, in: Toward Category-Level Object Recognition, Springer, Berlin, Heidelberg, 2006, pp. 382–400.
[93] C.H. Lampert, M.B. Blaschko, T. Hofmann, Beyond sliding windows: object localization by efficient subwindow search, in: IEEE Conference on Computer Vision and Pattern Recognition, 2008, CVPR 2008, IEEE, 2008, pp. 1–8.
[94] J.M. Stellman, Encyclopaedia of occupational health and safety, International Labour Organization, 1998.
[95] J. Irizarry, M. Gheisari, B.N. Walker, Usability assessment of drone technology as safety inspection tools, J. Inform. Technol. Constr. (ITcon) 17 (2012) 194–212.
[96] C.P. Papageorgiou, M. Oren, T. Poggio, A general framework for object detection, in: Sixth International Conference on Computer vision, 1998, IEEE, 1998, pp. 555–562.

Please cite this article in press as: J. Seo et al., Computer vision techniques for construction safety and health monitoring, Adv. Eng. Informat. (2015), http://dx.doi.org/10.1016/j.aei.2015.02.001