Advanced Engineering Informatics 45 (2020) 101100

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Real-time smart video surveillance to manage safety: A case study of a transport mega-project

Hanbin Luo[a,b], Jiajing Liu[a,b], Weili Fang[a,b,c,*], Peter E.D. Love[c], Qunzhou Yu[a,b], Zhenchuan Lu[a,b]

[a] Department of Construction Management, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
[b] Hubei Engineering Research Centre for Virtual, Safe, and Automated (VISAC), Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
[c] School of Civil and Mechanical Engineering, Curtin University, Perth, Western Australia, Australia

ARTICLE INFO

Keywords:
Computer vision
Construction safety
Video surveillance
Stuck-by accident

ABSTRACT

There is a tendency for accidents and even fatalities to arise when people enter hazardous work areas during the construction of projects in urban areas. A limited amount of research has been devoted to developing vision-based proximity warning systems that can determine when people enter a hazardous area automatically. Such systems, however, are unable to identify specific hazards and the status of a piece of plant (e.g., excavator) in real-time. In this paper, we address this limitation and develop a real-time smart video surveillance system that can detect people and the status of plant (i.e. moving or stationary) in a hazardous area. The application of this approach is demonstrated during the construction of a mega-project, the Wuhan Rail Transit System in China. We reveal that our combination of computer vision and deep learning can accurately recognize people in a hazardous work area in real-time during the construction of transport projects. Our developed systems can provide instant feedback concerning unsafe behavior and thus enable appropriate actions to be put in place to prevent their re-occurrence.

1. Introduction

Worldwide construction is a high-risk activity and a significant contributor to workplace accidents and fatalities that materialize from all industrial sectors [21,28,31,47]. During construction, however, it has been found that 25% of fatalities are associated with plant (e.g., backhoes and excavators) and equipment [3]. Moreover, 75% of the fatalities that occur are due to people coming into direct contact with plant and equipment [49].

Computer vision has been advocated to support behavior-based safety (BBS) to manage safety proactively. In this instance, it has been used to identify the conditions that can result in unsafe behaviors and then put in place mechanisms to prevent their occurrence [4,10,43,56,37]. A limited amount of research, however, has been undertaken that has examined the nature of accidents that arise when a piece of plant strikes an individual via monocular images or unmanned aerial vehicle (UAV) [48,19,20,23]. Despite the success of the work that has been undertaken to identify the unsafe behavior that has resulted in a person being struck by plant [19,20,23] several limitations that have hindered their accuracy, which include:

- The use of UAV's need to be manually operated. The transformation of each image's frame is also a manual process requiring the reference object (i.e. of known size) it be modified when outside the UAV's view [19,20,23];
- The UAV is unable to identify and process a person being stuck-by a piece of plant in real-time when an accident occurs [19,20,23];
- The plant's status (i.e. moving or stationary) has not to be considered by Kim et al. [19,20,23] and Son et al. [48], which may lead to error warning. For example, if a plant is not moving, it would be safe if a worker is approaching its area [19,20,23,48];
- The installation of a camera on a piece of plant has a limited field view. In particular, a camera is unable to be used in mega projects where multiple pieces of plant may be in operation [48]. As a result, this renders it impossible to detect plant and people who may be nearby simultaneously.

To address the above limitations, we develop a robust and efficient real-time smart video surveillance system to recognize when people enter a hazardous area. In comparison with other image devices (e.g., monocular and UAV), video surveillance can provide a more comprehensive view of sites, and can be used to monitor activities 24-hours a day in real-time. It follows, however, that the larger the area to be monitored, the more video cameras are required, which increases the

* Corresponding author at: Department of Construction Management, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China.
E-mail address: weili.fang@curtin.edu.au (W. Fang).

https://doi.org/10.1016/j.aei.2020.101100
Received 6 October 2019; Received in revised form 17 December 2019; Accepted 12 April 2020
Available online 18 May 2020
1474-0346/

complexity of computation.

In this paper, we present a novel real-time smart video surveillance system that can be used to detect plant and people automatically and can be used to contribute to preventing accidents on sites actively. We design our real-time smart video surveillance system draws on previous studies that have utilized computer vision and deep learning-based approaches to determine a piece of plant's status, and its proximity to people. To validate the effectiveness and feasibility of our developed real-time smart video surveillance system, we demonstrate its application during the construction of the Rail Transit System in Wuhan, China. We commence our paper by providing a review of the developments in proximity warning approaches. Our proposed research approach and case study are then presented. Next, we discuss our results and identify our studies, limitations. We conclude our paper by summarizing our work and identify avenues for future research.

2. Literature review

Computer vision has been used for a variety of purposes such as tracking vehicles and surveillance [5], traffic sign *retro*-reflectivity condition [1], and the inspection of rail tracks [57]. Deep learning has also been extensively used to predict issues associated with traffic flows [38], accidents from social media data [55], and bicycle demand forecasts [51]. In this paper, we focus on integrating computer vision with deep learning to detect the proximity of plant and people to one another. We will now briefly review developments in proximity warning approaches, object detection, and distance measurement.

2.1. Proximity warning approaches

The literature is replete with studies that have developed non-visual technology-enabled systems to detect people when they are near a piece of plant. For example, non-visual sensing techniques based on technologies such as Radio Frequency Identification (RFID) [26], Global Positioning System (GPS) [39], and Bluetooth [33] have been utilized to acquire location information to alert and inform people that they are too close to plant. In this instance, non-visual sensor devices are required to be attached to people and plant to enable the detection of their location in real-time.

Table 1 provides a summary of non-visual-based proximity warning systems studies. Despite the widespread use of non-visual sensors, signal attenuation is a significant challenge when using a locating sensor-based proximity warning system during construction. For example, GPS is ineffective within indoor environments as it provides inaccurate results when used in confined spaces. A major limitation of non-visual sensors is peoples unwilling to use them [22,54]. Thus, there is a need to develop a non-intrusive approach that can determine the location and a person's proximity to the plant in hazardous work areas.

The advent of high-resolution video cameras, the augmented storage capacity of databases and increasing accessibility of the Internet, have transformed the ability to document operations in construction and during an asset's maintenance. Consequently, computer vision-based applications have been developed to obtain information (e.g., locations, and actions) from digital images that can be automatically extracted and used for progress monitoring [15,52] defect detection [14,24] and identify unsafe behavior [6,7,9].

There has been a limited amount of research that has used computer vision to determine a person's proximity to plant [19,20,23,48]. The most important study that has been undertaken in this area that of Kim et al. [20] who integrated computer vision with a fuzzy inference module to create an automatic monitoring system. The developed system can determine safety levels on-site based on accidents that had occurred from moving objects. Kim et al. [20] used a Gaussian mixture model (GMM) to detect excavators and people by distinguishing between moving objects from their background. Then, the proximity between people and excavators were derived from their pixel size. A person's safety level was determined by inputting proximity and crowdedness obtained from a computer vision module, which was then inputted into one based on fuzzy interference. Despite the system's potential, Kim et al. [20] acknowledged that several improvements were required if their approach was to be applied during construction to enable real-time detection. Improvements needed are: (1) consideration of the plant's operational status; and (2) achieving a more significant level accuracy when dealing with high dimensional image data where there is a presence of clutter, numerous resources (e.g., people, and plant), varying poses and differing scales.

Similarly, Kim et al. [19] combined object detection (i.e., You Only Look Once (YOLOv3, 2018)) with distance measurement and a UAV to prevent people from being struck by a piece of plant. As we pointed out above in the introduction to our paper, there are limitations to this work. Son et al. [48] made headway in combating the limitations we identified by creating a vision-based proximity warning system that relies on the cameras mounted on the four sides of the plant. However, it falls short as it is unable to cater for larger-scale sites.

2.2. Object detection

Table 2 presents a summary of the extant literature that has examined object detection in construction. We can see from Table 2 that several computer vision-based approaches have been developed to detect objects (i.e., people, plant, materials) from images. Deep learning incorporating Convolutional Neural Networks (CNN) have been demonstrated to be an effective method for object detection [12,40,42]. Object detection comprises two approaches: (1) handcrafted feature-based; and (2) deep learning-based. For example, Memarzadeh et al. [30] developed a computer vision approach using Histogram of Oriented Gradients (HOG) and Hue Saturation Value (HSV) descriptors to detect individuals and excavators. Similarly, Fang et al. [8] applied Faster R-CNN approach to detect people and heavy equipment, which achieved a better performance than other CNN approaches.

The works presented in Table 2 have demonstrated that they can successfully detect objects. With the developments of deep learning,

Table 1
Non-visual-based proximity warning systems studies.

Technology	Description	Limitation	Authors
Bluetooth	Consideration of equipment types and approaching speeds to avoid system delays	Take more time on debugging device	Park et al. [34]
Chirp Spread Spectrum (CSS)	Consideration of temporal and spatial factors to assess safety risks	The multipath effect has a significant effect on wireless ranging CSS signals	Luo et al. [29]
GPS	Generating heat maps for visual safety management	Be subjective on the determination of influence degree	Golovina et al. [13]
Magnetic field	Acquiring a person's exact location relative to specific parts of the mining machine based on received magnetic flux density	The prototype system is limited to applications in laboratory environments	Li et al. [27]
RFID	Providing blind spot information for equipment operators to avoid potential hazard due to limited visibility	Signal propagation is difficult to maintain reliable performance in changing environments	Teizer et al. [50]

Table 2
A summary of prior works on computer vision-based approaches for object detection.

Categories	Descriptions	Authors
Handcrafted feature-based	HOG descriptor to detect hydraulic excavators	Azar and Mccabe [2]
	Applied background subtraction, HOG, HSV descriptor and *k*-NN classifier to detect individuals	Park and Brilakis [35]
	Applied background subtraction and HOG descriptor, and support vector machine to detect individuals and their hardhat	Park et al. [36]
Deep learning-based	Faster R-CNN to detect people and heavy equipment	Fang et al. [7]
	Mask R-CNN to detect people and structural supports	Fang et al. [9]
	YOLOv3 to detect people and plant	Kim et al. [19]

several CNN-based object detection approaches have emerged from the field of computer science, which have provided new opportunities for accurately detecting objects. The Faster R-CNN, YOLO, and Single Shot Multibox Detector (SSD) have been the most commonly applied for detecting objects [40]. In comparison with the Faster R-CNN (17 frames per second (FPS)), SSD (16FPS), and YOLOv3(35FPS), YOLOv2 (40FPS) have attained faster performance levels for Microsoft's Common Objects in context (MS COCO) dataset [41]. With this in mind, YOLOv2 is selected as the detection approach in this research due to its ability to quickly identify objects.

2.3. Distance measurement

Several approaches have been developed to measure distances between objects. For example, three-dimensional (3D) sensing devices (e.g., stereo-vision camera, depth sensor (i.e., Kinect)) have been used to determine the distance from given 3D spatial information. However, such sensing devices are limited in their range and are sensitive to lighting, and therefore, they are unable to be used in construction [19]. For example, Kinect devices have a limited range of four meters and are extremely sensitive to light [16].

In addressing these limitations of computing distance from a single camera, Kim et al. [19,20,23] have proposed several solutions. For example, Kim et al. [23] applied a transformation matrix to represent the geometric relationship among objects. Here the distance between objects is estimated by measuring pixel distance between them where a reference object's geometric is known. Despite the success of this approach, it is unable to measure the distance between objects as it becomes inaccurate with losses of in-depth information. It has been suggested that a 3D reconstruction can be used to compute its distance to improve its accuracy [11,25,53]. However, 3D reconstruction approaches require considerable computational power, which can render its application for real-time detection objects to be inappropriate.

3. Research approach

The approach that we adopted to design and develop our computer vision and deep learning model consists of four procedures: (1) data collection; (2) detection of objects; (3) recognition of objects status; and (4) recognition of people in hazardous areas. We present the workflow used to develop our approach in Fig. 1 and describe each of the procedures performed below.

3.1. Data collection of images for training

To utilize the benefits that can be enabled by computer vision, there is a need to obtain video surveillance during the construction process. Databases that can be used to train deep learning models for object detection during the construction of mega transport projects are not readily available. Thus, there was a need to create a database to train the model that we developed and test its ability to detect objects. To avoid bias, we collected different views and aspects of images (e.g., scale, and illuminations). The database was randomly divided into two parts: (1) training; and (2) testing.

After creating the database, we used the LabellmgR [32] annotation tool, written in Python and the Qt to develop a graphical interface to enable the manual labeling of images. The format of each label was saved in an "XML file" format. To apply our computer vision approach, we needed to have access to video stream from surveillance in real-time. We developed a program code written in C++ to acquire the video data and assess the video surveillance using the Software Development Kit (SDK) interface.

3.2. YOLOv2-based object detection

YOLOv2 is a deep CNN that performs real-time object detection for 80 object classes. Compared with the Faster R-CNN and SSD, YOLOv2 is the most efficient real-time object detection approach, which has a higher recognition rate and processing speed. YOLOv2 is an improved version of YOLOv1, which was proposed by Redmon and Farhadi [40]. The Darknet-19 YOLOv2 architecture and parameters are aligned with the default setting presented in Redmon and Farhadi [40]. We present a description of the critical aspects of YOLOv2 below. We adopt the Darknet-19 YOLOv2 model and use our database to fine-tune it to detect objects on construction sites (e.g., people and excavators).

YOLOv2 adopts the anchor ideas of the Faster R-CNN to improve its accuracy to detect small objects, which samples on the convolution feature map with a sliding window. YOLOv2 extracts features for an entire input image using 19 convolutional layers and five max-pooling layers to jointly predict an object's location (bounding boxes) and the object's confidence scores for 361 separate grid cells. We refer readers to Redmon and Farhadi [40] for further information about YOLOv2.

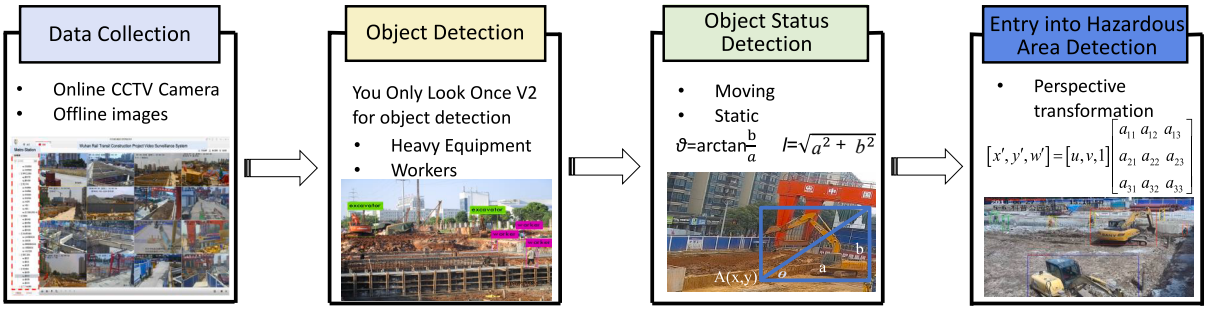

Fig. 1. The workflow of our proposed computer vision approach.

For each cell, YOLOv2 provides predictions for five different boxes. For each box, the prediction contains its confidence (i.e., the probability of the box containing an object), location, and the probability of each class label. A box is discarded when the product's confidence and the probability of the most likely class are below a threshold value (e.g., 0.3 in our experiments). Finally, a non-max suppression algorithm is applied in a post-processing phase to discard redundant boxes that significantly overlap [44].

3.3. Recognition of plant status

Several approaches have been used to detect moving objects such as the background subtraction [17] and optical flow methods [45,46]. Such approaches are, however, prone to experiencing delays in detection due to issues surrounding their processing speed. To ensure our computer vision approach can detect people in a hazardous work area in real-time from a video stream, we proposed a new pipeline by using the coordinates of the object that are obtained from YOLOv2 to detect a plant's status during operations.

In this pipeline, two key parameters determine whether the plant is stationary or moving, which are the: (1) coordinates of the rectangular lower left vertex; and (2) angle with the horizontal direction. A bounding box is obtained when detecting the plant, which in the example that we present is an excavator (Fig. 2). The two key parameters can be calculated using Eqs. (1) and (2), respectively. If the values of "Δd" and "Δθ" in a continuous frame are unchanged or minor changed, the excavator would be regarded as being stationary.

$$\Delta d = \sqrt{(x' - x)^2 + (y' - y)^2} \tag{1}$$

$$\Delta\theta = \arctan\frac{b}{a} - \arctan\frac{b'}{a'} \tag{2}$$

3.4. Detecting people in hazardous work area

The final step for recognizing a person in a hazardous work area is to derive their proximity from the plant using camera surveillance. A reverse perspective is used in this research to derive this proximity. The transformation process required four steps: (1) access to the camera's internal parameters and a distortion model; (2) acquisition of the pixel coordinates of the four known points in the image; (3) calculation of the transformation matrix; and (4) the generation of the transformed image. Eq. (3) presents the transformation perspective, which indicates the coordinate relationship of points in the original image and the new viewing plane.

$$[x', y', w'] = [u, v, 1]\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{3}$$

where $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ are the transformation matrix to implement linear (such as scaling, shearing or rotation), translation, and perspective transformation operations, and $a_{(11\text{-}33)}$ stand for the transformation coefficients; $[u, v]$ are the coordinates of the point in the original image; $[x', y']$ refers to the rectified coordinates of an original pixel; and w' stands for a scale factor. Eq. (3) can be further derived into Eq. (4):

$$\begin{cases} x = \frac{x'}{w'} = \frac{a_{11}u + a_{21}v + a_{31}}{a_{13}u + a_{23}v + a_{33}} \\ y = \frac{y'}{w'} = \frac{a_{12}u + a_{22}v + a_{32}}{a_{13}u + a_{23}v + a_{33}} \end{cases} \tag{4}$$

where (x, y) is the new coordinate of the point on the projection plane.

As can be seen in Eq. (4), the corresponding coordinate relationships of the four arbitrary points between the original and new image form the basis for calculating the transformation matrix. We present an example of the reverse perspective in Fig. 3. A referenced object (i.e., its known size) computes the size of a pixel in an image. In doing this, the Euclidean geometric distance between two objects is calculated. Regardless of the plant's state, the system can determine if an alarm needs to be issued if a person is in close proximity (see Fig. 4).

The first step of using computer vision is to recognize people in a hazardous work area from a video stream and define a series of safety rules. In this instance, we draw on the work of Shen et al. [48] who have derived safety rules for determining hazardous areas for plant during construction. The excavator was selected as the plant of interest in this experiment due to its widespread use on the construction site and the associated high accident rate [18]. For excavator, three warning levels are predefined: (1) red warning (0–7.7 m); (2) yellow warning (7.7–16 m); and (3) safe condition (> 16 m).

All the algorithms were executed on a server with Intel(R) Xeon(R) CPU, 64 GB of RAM, NVIDIA Quadro p5000 of the graphics card.

4. Case study

The Wuhan Rail Transit System (Hubei province China) is an elevated and underground system. Working in collaboration with a contractor, more than 240 security Hikvision cameras (i.e., DS-2DF8223IW-A) with two million pixel high-definition were installed on

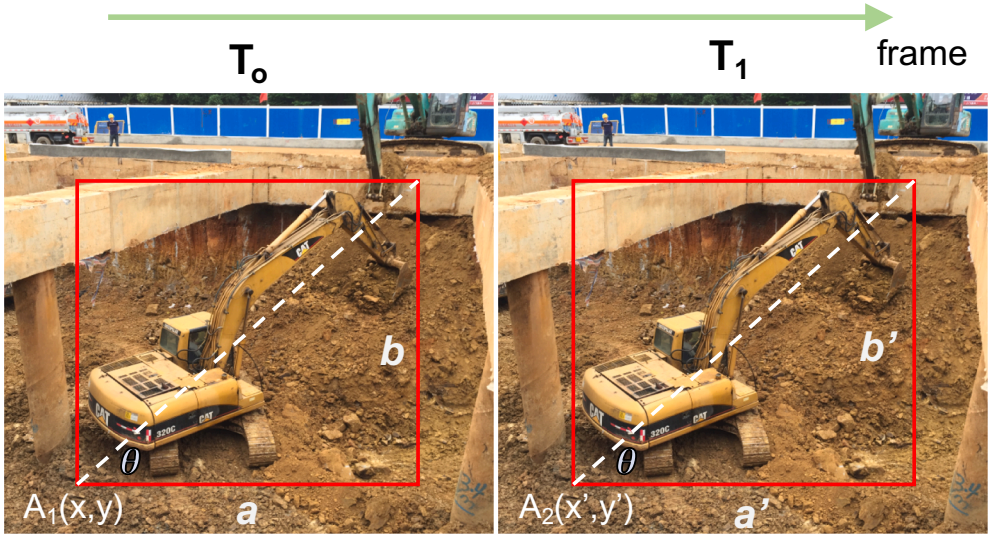

Fig. 2. An example of the excavator's bounding box.

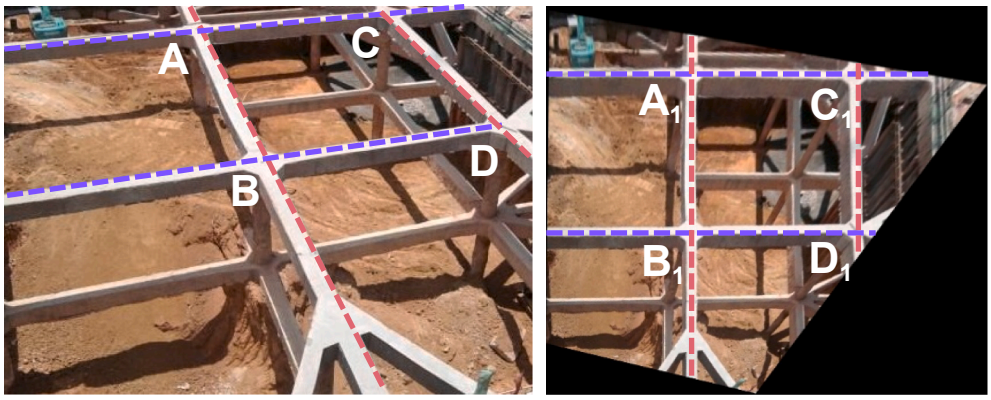

Fig. 3. Use of a reserve perspective transformation on a referenced object.

several sites and used to record daily activities in real-time during construction (Fig. 5).

4.1. Data collection

Two types of image data were collected from the selected construction site to: (1) create a database of photographs of people and plant for training YOLOv2 model; and (2) assess the surveillance's SDK to recognize when people were in a hazardous area in real-time.

4.1.1. Creation of YOLOv2 training database

As mentioned above, people and excavators in the dataset were collected from different viewpoints and aspects (e.g., varying scales, poses, occlusions, and lighting conditions) (Fig. 6). In total, more than 10,000 labeled image frames containing people and excavators were obtained. All of these labeled images were used to train our YOLOv2 model to test its performance to detect objects. A total of 1000 pieces of data were randomly selected for assessment.

4.1.2. Assessing surveillance SDK for testing in real-time

To enable our computer vision approach to assess the video surveillance obtained and recognize a hazardous area in real-time, we initially needed to assess Hikvision's SDK using the following steps:

(1) Initiate of Hikvision's SDK system and pre-allocate the memory.
(2) Set connection timeout and a callback function to receive exception messages.
(3) Register user. Predefine a device parameter structure and pass the device's IP password port to it, and then call the registration function for user registration.
(4) Start preview. Define a configuration structure for the Internet Protocol (IP) device resource and channel resource. Then obtain the device configuration information. Therefore, the information of the IP channel can be directly obtained from the device, and the preview module of the parameter structure can be invoked through the defined corresponding channel number for real-time preview.
(5) Stop preview, logout, and release the SDK resources.

4.2. Detection of people and excavators

The training of the YOLOv2 model consisted of two parts: (1) classification; and (2) detection. In this research, the parameters for training the YOLOv2 model were derived from Redmon and Farhadi [40]. The classification model of the Darknet-19 was trained using our created database of objects (i.e., people and excavators) for 160 epochs with a stochastic gradient descent with a starting learning rate of 0.001, polynomial rate decay with a power of 4, a weight decay of 0.0005 and momentum of 0.9. The detection model is also trained for 160 epochs with a starting learning rate of 0.001 and a tenfold reduction at 60 and 90 epochs.

Two key performance indicators (KPI) were used to measure the effectiveness of the YOLOv2-based object detection approach: (1) precision; and (2) recall. The value of precision and recall rates can be calculated using the following Eq. (5):

$$\begin{cases} \mathrm{Pr}\,ecision = \frac{TP}{TP+FP} \\ \mathrm{Re}\,call = \frac{TP}{TP+FN} \end{cases} \tag{5}$$

where FP refers to the mistaken identity of other objects like people or excavators, FN refers to the undetected people and equipment, and TP

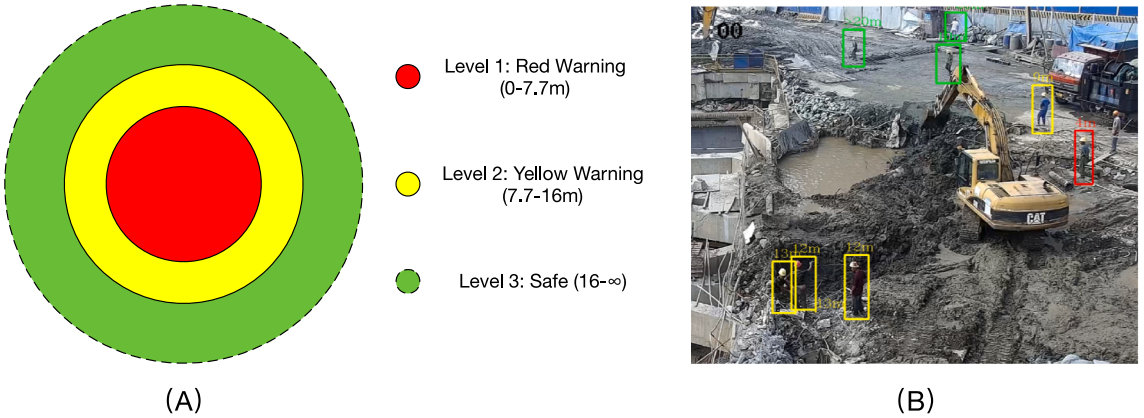

Fig. 4. Examples of warning levels.

Fig. 5. A web-based real-time monitoring system.

refers to people and equipment correctly detected. In this research, a threshold value of 'True Positive' is set to 0.3. Table 3 presents the testing results of using YOLOv2 to detect people and excavators. It also demonstrates that our YOLOv2 model can achieve a high level of accuracy in detecting people and excavators during construction. Fig. 7 provides an example of our detection results. However, factors such as occlusions and an object's scale affect the accuracy of our YOLOv2 approach to detect people and excavators.

Table 3
People and excavator detection results with a testing dataset (detection with p (object|box > 0.3).

Objects	Correctly Detected objects (TP)	Incorrectly Detected Objects		Precision	Recall
		Not detected (FN)	Mis-detected (FP)		
People	1896	309	125	94%	86%
Excavator	843	107	90	90%	89%

4.3. Recognition of excavator status

To validate the effectiveness of our approach, we were required to

Fig. 6. Examples of images in the dataset.

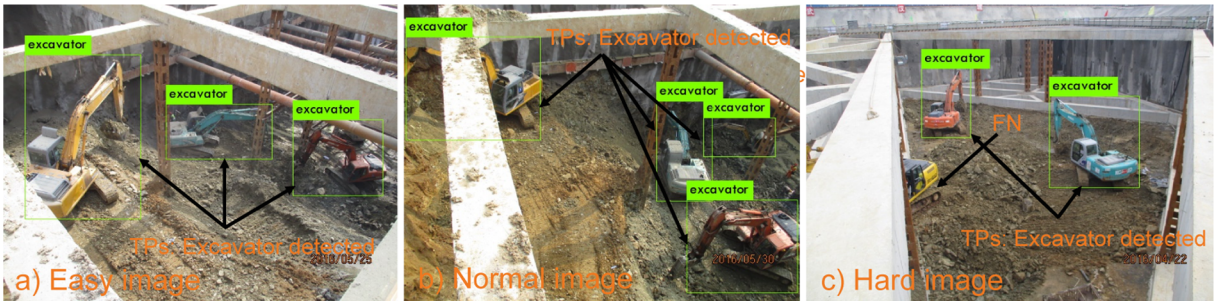

Fig. 7. Examples of detecting people and excavators from different perspectives.

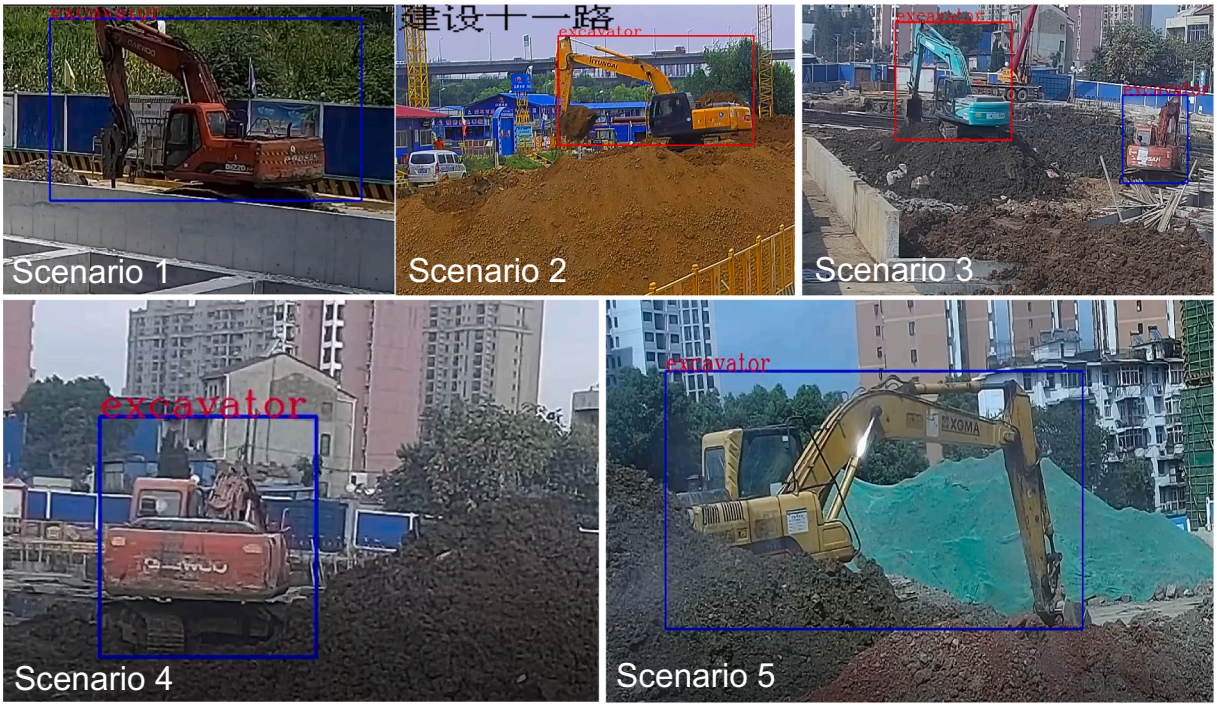

Fig. 8. Five construction scenarios for recognition of excavators' status.

recognize the status of excavators. Therefore, we select five different scenarios from our real-time video surveillance system (Fig. 8). In this research, the surveillance system collected images with 25 FPS. If all frames were inputted into our approach to detect people in a hazardous area, then a time delay in computation speed would be experienced due to its low speed. To reduce the computation time, we, therefore, set a detection rate to be one FPS in our system. As mentioned above, the status of an excavator (i.e., moving (red bounding box) or static (blue bounding box)) was determined by two key parameters: (1) the coordinates of the detection bounding box (Δd); (2) and the direction of the bounding box's diagonal (Δθ). Here, a threshold value of "Δθ" is set to 5°; and "Δd" is set to 10 pixels.

In Table 4, we present the details of the selected videos and detection scenario results. We reveal that our approach was able to recognize the status of excavator with an average accuracy of 91%.

4.4. Recognition of people in hazardous areas

After correctly detecting and recognizing objects (i.e., people and excavators), we then sought to derive the proximity between them

Table 4
Detection results of recognizing the status of excavator.

Scenario	Status of Excavator	Ground Truth		Detection Results	Average Accuracy
		Stationary	Moving		
1	Stationary status	300	0	275 correctly detected (stationary) and 25 incorrectly detected (moving)	91%
2	Moving status	0	300	279 correctly detected (moving) and 21 incorrectly detected (stationary)	
3	Two excavators: Stationary status (A) and moving status (B)	300	300	Excavator A: 265 correctly detected and 35 incorrectly; Excavator B: 284 correctly detected and 16 incorrectly detected	
4	Changing status (stationary (A) to moving (B))	138	162	Status A: 123 correctly detected and 15 incorrectly; status B: 146 correctly detected and 16 incorrectly detected	
5	Changing status (moving (A) to static (B))	117	183	Status A: 115 correctly detected and two incorrectly; status B: 157 correctly detected and 26 incorrectly detected	

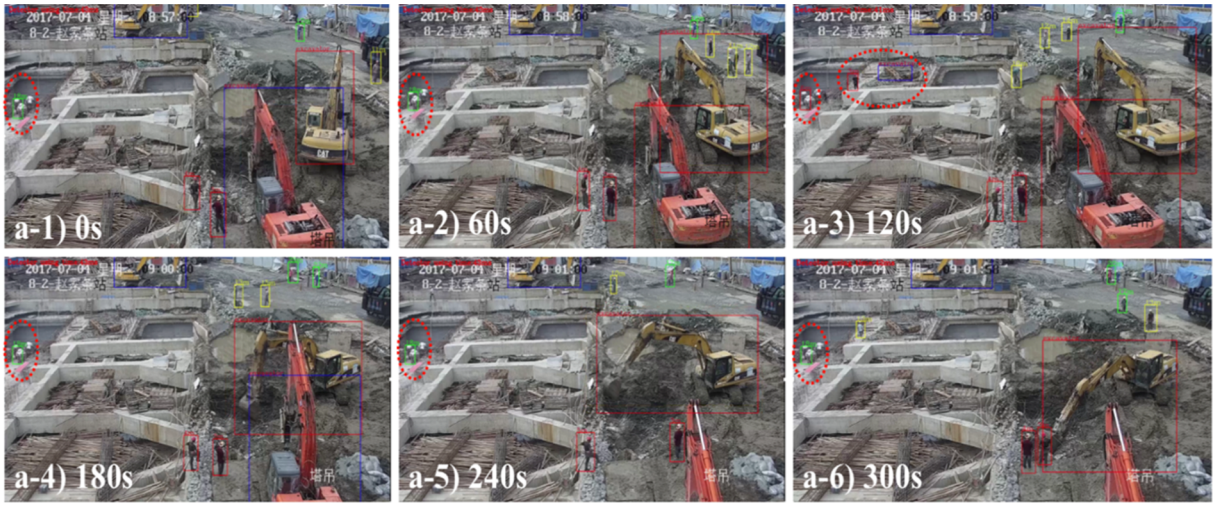

Fig. 9. Detection results for a hazardous are from video surveillance.

using two case examples.

In the first example we present, a beam is taken as a referenced object to compute the representation size of each pixel (Fig. 3). Then, the proximity between a person and excavator is determined by using the reserve perspective transformation, as highlighted previously. The typical length of a deep foundation pit (L_{AC}) was selected in the project as a referenced object (Fig. 3). Fig. 9 presents a snapshot of the recognition results.

While our approach was successfully able to recognize people in hazardous areas from video surveillance, we need to acknowledge that several misdetection and errors were made. In Fig. 9, we identify in red an example of an error detection. The main reason is that our object detection approach is unable to detect people and excavators simultaneously accurately.

In this second example, Fig. 10 presents examples of the detection results in red (0–150 s).

5. Discussion

We developed a real-time smart video surveillance monitoring system that integrated techniques from computer vision and deep learning to recognize when a person entered a hazardous area using video. When a person enters a hazardous area, an alert is generated, enabling the likelihood of the person being struck by plant to be reduced. Our developed system not only enables site management to identify when an unsafe act has occurred but also records the action in real-time. This information can be used to demonstrate to the culprit that actions were unsafe, and an accident may have occurred. Based on the proximity between people and excavator, our proposed system can visualize the safety levels of people that are in proximity to the excavator.

The contributions of the research presented in our paper are twofold. First, we have developed a robust method that combines computer vision and Yolov2 to recognize people in a hazardous area from video streams in real-time. We have also been able to detect when an excavator is stationary or moving, which previous research has been unable to do. Second, in comparison with the COCO and ImageNet databases, the database created and used in this research is adequate for the characteristics specific to construction (i.e., cluttered with objects, dynamic and complex condition). It takes a long time to extract images features from video streams using YOLOv2 (23FPS), even though YOLOv2 has achieved a high speed at 67FPS in ImageNet database [40]. Our strategy adapted one frame/s of video surveillance as input enabling recognition of people entering hazardous areas from a video stream in real-time.

6. Limitations

We need to acknowledge the limitations of our research. Our study lacked an effective measure to validate the precision of the distance between people and the plant (i.e. excavator) we examined. The distance between people and the excavator was obtained using a reserve

Fig. 10. Examples of detection results.

perspective transformation approach from a two-dimensional (2D) camera. However, improvements in the precision of distance between people and excavator can be made by producing a 3D model to extract information from a stereo camera. Moreover, a validation experiment should be taken by combining non-visual sensors (e.g., GPS) to enable a sound and reliable monitoring system to be developed and implemented in real-time.

The safety rules adapted from Shen et al. [48] need to be extended to enable a valid and reliable system to be implemented. In this paper, the excavator's operational speed was assumed to be constant; however, it will no doubt vary during construction. The excavator's speed and operation direction would be directly exploited from successive frames to address this problem in our future research.

Cluttered construction sites can hinder the ability to recognize people and plant, which influences the accuracy to recognize objects in hazardous areas. Furthermore, the proximity of people and excavators were derived from a single image by using a reserve perspective transformation. By using this approach to derive proximity, an object (i.e., of known size) should be used to compute the representation size of each pixel. However, this was not the case as the precision of distance could not be accurately compared with the actual proximity that existed between people and excavators. We also assumed that the cameras positions, view, and zoom remained constant. These factors will affect the distance measurement performance. For example, if the view changes, the computing process would be subject to modification. Finally, for our proposed approach to be applied in practice, the size of the database needs to be increased to accommodate varying different poses, scales, and construction sites, to improve its accuracy.

7. Conclusions

In this paper, we have presented a real-time smart video surveillance monitoring system to recognize people entering dangerous areas on a construction site. Three algorithms were developed: (1) a YOLOv2-based detection approach is used to detect objects; (2) a technique is used to determine the status of the heavy plant; and (3) a reserve perspective transformation to derive proximity between people and plant. A mega transport project was used to validate the effectiveness and feasibility of our smart video surveillance system during its construction. The results demonstrated that our system was able to recognize objects accurately; in this case, people and excavators in hazardous areas from video surveillance in real-time.

Considering the accuracy of our developed system, we suggest that there is potential for it to be implemented in practice. Though, this will require a considerably more extensive database than the one developed for this study. Being able to recognize people's unsafe behavior in real-time from video surveillance can result in effective managerial interventions and immediate behavior modification. Moreover, the detected videos not only can be used to provide individual's with direct visual feedback about their unsafe actions but also as a tool for safety education.

To enact real-time monitoring, on-site and improve safety performance will require further research, involving a stereo camera to be installed on construction sites. We suggest that by collecting data in a 3D format from the stereo camera, spatial information can be obtained, which can enable accurate measurement of the distance between objects.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research is supported by National Natural Science Foundation of China (Grant No.71732001, No.51978302, No.51678265, No. 51878311).

References

[1] C. Ai, Y. Tsai, An automated sign retroreflectivity condition evaluation methodology using mobile LIDAR and computer vision, Transport. Res. C-Emer. 63 (2016) 96–113.
[2] E.R. 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.
[3] J.E. Beavers, J.R. Moore, R. Rinehart, W.R. Schriver, Crane-related fatalities in the construction industry, J. Constr. Eng. M. 132 (9) (2006) 901–910.
[4] P. Cocca, F. Marciano, M. Alberti, Video surveillance systems to enhance occupational safety: A case study, Safety. Sci. 84 (2016) 140–148.
[5] B. Coifman, D. Beymer, P. McLauchlan, J. Malik, A real-time computer vision system for vehicle tracking and traffic surveillance, Transport. Res. C: Emer. 6 (4) (1998) 271–288.
[6] L. Ding, W. Fang, H. Luo, P.E.D. Love, B. Zhong, X. Ouyang, A deep hybrid learning model to detect unsafe behavior: Integrating convolution neural networks and long short-term memory, Automat. Constr. 86 (2018) 118–124.
[7] W. Fang, L. Ding, H. Luo, P.E.D. Love, Falls from heights: A computer vision-based approach for safety harness detection, Automat. Constr. 91 (2018) 53–61.
[8] W. Fang, L. Ding, B. Zhong, P.E.D. Love, H. Luo, Automated detection of workers and heavy equipment on construction sites: A convolutional neural network approach, Adv. Eng. Inform. 37 (2018) 139–149.
[9] W. Fang, B. Zhong, N. Zhao, P.E. Love, H. Luo, J. Xue, S. Xu, A deep learning-based approach for mitigating falls from height with computer vision: Convolutional neural network, Adv. Eng. Inform. 39 (2019) 170–177.
[10] W. Fang, P.E.D. Love, H. Luo, L. Ding, Computer vision for behaviour-based safety in construction: A review and future directions, Adv. Eng. Inform. 43 (2020) 100980.
[11] H. Fathi, I. Brilakis, Multistep Explicit Stereo Camera Calibration Approach to Improve Euclidean Accuracy of Large-Scale 3D Reconstruction, J. Comput. Civil. Eng. 30 (1) (2016) 04014120.
[12] R. Girshick, Fast R-CNN, Ieee. I. Conf. Comp. Vis. (2015) 1440–1448.
[13] Golovina, J. Teizer, N. Pradhananga, Heat map generation for predictive safety planning: Preventing struck-by and near miss interactions between workers-on-foot and construction equipment, Automat. Constr. 71 (2016) 99–115.
[14] K. Gopalakrishnan, S.K. Khaitan, A. Choudhary, A. Agrawal, Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection, Constr. Build. Mater. 157 (2017) 322–330.
[15] K.K. Han, M. Golparvar-Fard, Appearance-based material classification for monitoring of operation-level construction progress using 4D BIM and site photologs, Automat. Constr. 53 (2015) 44–57.
[16] S. Han, S. Lee, A vision-based motion capture and recognition framework for behavior-based safety management, Automat. Constr. 35 (2013) 131–141.
[17] S.R. Hanchinamani, S. Sarkar, S.S. Bhairannawar, Design and Implementation of High Speed Background Subtraction Algorithm for Moving Object Detection, Procedia. Comput. Sci. 93 (2016) 367–374.
[18] J.W. Hinze, J. Teizer, Visibility-related fatalities related to construction equipment, Safety. Sci. 49 (5) (2011) 709–718.
[19] D. Kim, M.Y. Liu, S. Lee, V.R. Kamat, Remote proximity proximity monitoring between mobile construction resources using camera-mounted UAVs, Automat. Constr. 99 (2019) 168–182.
[20] H. Kim, K. Kim, H. Kim, Vision-Based Object-Centric Safety Assessment Using Fuzzy Inference: Monitoring Struck-By Accidents with Moving Objects, J. Comput. Civil. Eng. 30 (4) (2016) 04015075.
[21] S. Guo, L. Ding, Y. Zhang, M.J. Skibniewski, K. Liang, Hybrid recommendation approach for behavior modification in the Chinese construction industry, J. Constr. Eng. Manage. 145 (6) (2019), https://doi.org/10.1061/(ASCE)CO.1943-7862.0001665.
[22] J. Kim, S. Chi, J. Seo, Interaction analysis for vision-based activity identification of earthmoving excavators and dump trucks, Automat. Constr. 87 (2018) 297–308.
[23] K. Kim, H. Kim, H. Kim, Image-based construction hazard avoidance system using augmented reality in wearable device, Automat. Constr. 83 (2017) 390–403.
[24] C. Koch, K. Doycheva, V. Kasireddy, B. Akinci, P. Fieguth, A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure, Adv. Eng. Inform. 30 (2) (2015) 208–210.
[25] E. Konstantinou, I. Brilakis, Matching Construction Workers across Views for Automated 3D Vision Tracking On-Site, J. Constr. Eng. M. 144 (7) (2018) 04018061.
[26] H.S. Lee, K.P. Lee, M. Park, Y. Baek, S. Lee, RFID-Based Real-Time Locating System for Construction Safety Management, J. Comput. Civil. Eng. 26 (3) (2012) 366–377.
[27] J.C. Li, J. Carr, C. Jobes, A shell-based magnetic field model for magnetic proximity detection systems, Safety. Sci. 50 (3) (2012) 463–471.
[28] P.E.D. Love, P. Teo, J. Smith, F. Ackermann, Y. Zhou, The nature and severity of workplace injuries in construction: engendering operational benchmarking, Ergonomics. 62 (10) (2019) 1273–1288.
[29] X. Luo, H. Li, T. Huang, M. Skitmore, Quantifying Hazard Exposure Using Real-Time

Location Data of Construction Workforce and Equipment, J. Constr. Eng. M. 142 (8) (2016) 04016031.
[30] 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.
[31] S. Guo, J. Li, K. Liang, B. Tang, Improved safety checklist analysis approach using intelligent video surveillance in the construction industry: a case study, Int. J. Occup. Saf. Ergo. (2019), https://doi.org/10.1080/10803548.2019.1685781.
[32] Tzutalin. LabelImg. Git code. < https://github.com/tzutalin/labelImg > , 2015 (last accessed on 1 December 2019).
[33] J. Park, E. Marks, Y.K. Cho, W. Suryanto, Performance Test of Wireless Technologies for Personnel and Equipment Proximity Sensing in Work Zones, J. Constr. Eng. M. 142 (1) (2016) 04015049.
[34] J. Park, X.Y. Yang, Y.K. Cho, J. Seo, Improving dynamic proximity sensing and processing for smart work-zone safety, Automat. Constr. 84 (2017) 111–120.
[35] M.W. Park, I. Brilakis, Construction worker detection in video frames for initializing vision trackers, Automat. Constr. 28 (2012) 15–25.
[36] M.W. Park, N. Elsafty, Z.H. Zhu, Hardhat-Wearing Detection for Enhancing On-Site Safety of Construction Workers, J. Constr. Eng. M. 141 (9) (2015) 04015024.
[37] M.E. Mneymneh, M. Abbas, H. Khoury, Evaluation of computer vision techniques for automated hardhat detection in indoor construction safety applications, Front. Eng. Manage. (2018), https://doi.org/10.15302/j-fem-2018071.
[38] N.G. Polson, V.O. Sokolov, Deep learning for short-term traffic flow prediction, Transport. Res. C-Emer. 79 (2017) 1–17.
[39] N. Pradhananga, J. Teizer, Automatic spatio-temporal analysis of construction site equipment operations using GPS data, Automat. Constr. 29 (2013) 107–122.
[40] J. Redmon, A. Farhadi, YOLO9000: Better, Faster, Stronger, Proc. Cvpr. IEEE. (2017) 7263-7271.
[41] J. Redmon, A. Farhadi, Yolov3: An incremental improvement, arXiv preprint arXiv:1804.02767 (2018).
[42] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, Adcanceds Neural Inform. Process. Syst. (2015) 91–99.
[43] M. Roetting, Y.H. Huang, J.R. McDevitt, D. Melton, When technology tells you how you drive–truck drivers' attitudes towards feedback by technology, Transport. Res. Part F: Traff. 6 (4) (2003) 275–287.
[44] R. Rothe, M. Guillaumin, L. Van Gool, Non-maximum suppression for object detection by passing messages between windows, Asian conference on computer vision, Springer, 2014, pp. 290–306.
[45] S.S. Sengar, S. Mukhopadhyay, Detection of moving objects based on enhancement of optical flow, Optik. 145 (2017) 130–141.
[46] S.S. Sengar, S. Mukhopadhyay, Motion detection using block based bi-directional optical flow method, J. Visual Commun. Image Representation 49 (2017) 89–103.
[47] X. Shen, E. Marks, N. Pradhananga, T. Cheng, Hazardous Proximity Zone Design for Heavy Construction Excavation Equipment, J. Constr. Eng. M. 142 (6) (2016) 05016001.
[48] H. Son, H. Seong, H. Choi, C. Kim, Real-Time Vision-Based Warning System for Prevention of Collisions between Workers and Heavy Equipment, J. Comput. Civil. Eng. 33 (5) (2019) 04019029.
[49] Statistics, U.S.B.O.L., Census of fatal occupational injuries. U.S.bureau of Labor Statistics, 2015.
[50] 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.
[51] C. Xu, J. Ji, P. Liu, The station-free sharing bike demand forecasting with a deep learning approach and large-scale datasets, Transport. Res. C-Emer. 95 (2018) 47–60.
[52] J. Yang, M.-W. Park, P.A. Vela, M. Golparvar-Fard, Construction performance monitoring via still images, time-lapse photos, and video streams: Now, tomorrow, and the future, Adv. Eng. Inform. 29 (2) (2015) 211–224.
[53] M.D. Yang, C.F. Chao, K.S. Huang, L.Y. Lu, Y.P. Chen, Image-based 3D scene reconstruction and exploration in augmented reality, Automat. Constr. 33 (2013) 48–60.
[54] M. Zhang, T. Cao, X. Zhao, Applying Sensor-Based Technology to Improve Construction Safety Management, Sens.-Basel. 17 (8) (2017).
[55] Z. Zhang, Q. He, J. Gao, M. Ni, A deep learning approach for detecting traffic accidents from social media data, Transport. Res. C: Emer. 86 (2018) 580–596.
[56] W. Fang, L. Ding, P.E.D. Love, H. Luo, H. Li, F. Peña-Mora, B. Zhong, C. Zhou, Computer vision applicatios in construction safety assurance, Automat. Constr. 110 (2020) 103013.
[57] L. Zhuang, L. Wang, Z. Zhang, K.L. Tsui, Automated vision inspection of rail surface cracks: A double-layer data-driven framework, Transport. Res. C: Emer. 92 (2018) 258–277.