

Bahaa Eddine MNEYMNEH, Mohamad ABBAS, Hiam KHOURY

Evaluation of computer vision techniques for automated hardhat detection in indoor construction safety applications

© The Author(s) 2018. Published by Higher Education Press. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0>)

Abstract Construction is considered among the most dangerous industries and is responsible for a large portion of total worker fatalities. A construction worker has a probability of 1-in-200 of dying on the job during a 45-year career, mainly due to fires, falls, and being struck by or caught between objects. Hence, employers must ensure their workers wear personal protective equipment (PPE), in particular hardhats, if they are at risk of falling, being struck by falling objects, hitting their heads on static objects, or coming in proximity to electrical hazards. However, monitoring the presence and proper use of hardhats becomes inefficient when safety officers must survey large areas and a considerable number of workers. Using images captured from indoor jobsites, this paper evaluates existing computer vision techniques, namely object detection and color-based segmentation tools, used to rapidly detect if workers are wearing hardhats. Experiments are conducted and the results highlight the potential of cascade classifiers, in particular, to accurately, precisely, and rapidly detect hardhats under different scenarios and for repetitive runs, and the potential of color-based segmentation to eliminate false detections.

Keywords construction, safety, personal protective equipment, hardhat, computer vision

1 Background

Despite different provisions aimed to create a safer working environment, construction remains among the

most perilous industries, responsible for a large portion of total worker injuries, risks, and fatalities (Mneymneh et al., 2016; Abbas et al., 2016; Abbas et al., 2018). In 2014, the construction sector was responsible for 899 fatal injuries in the United States, second only to the trade and transportation sector with 1246 fatal injuries; the mining sector had 183 (United States Department of Labor, 2014). More specifically, it was found that fires, falls, and being struck by or caught between objects contribute to over 50% of the total casualties in the sector. Furthermore, a large portion of work-related head injuries, in particular, is typically sustained by workers not wearing hardhats (United States Department of Labor, 2014). Hence, the proper use and adoption of safety equipment such as personal protective equipment (PPE), in particular, hardhats, was deemed necessary on jobsites to reduce the risk of injury by impact from falling or flying objects (Health and Safety Executive, 2014; Shrestha et al., 2015; Occupational Safety and Health Administration, 2016). However, because the awareness and attitude of construction workers toward the importance of hardhats cannot be fully trusted, safety personnel are typically deployed on construction sites to ensure compliance with safety regulations and maintain acceptable working conditions. Nonetheless, the current monitoring of hardhat-wearing remains manual, tedious, and time-consuming (Gheisari et al., 2014; Ham et al., 2016). Therefore, there is a significant requirement to automate this process in a cost-effective manner with highly reduced turnaround times to mitigate the risks associated with hazardous situations.

Among several information technology (IT) and computer-based tools widely adopted in the construction field to automate different processes (Kim et al., 2009; Khoury and Kamat, 2009a, 2009b; Kopp et al., 2010; Chae and Yoshida, 2010; Chdid et al., 2011; Khoury et al., 2012; Oueiss et al., 2012; Ding et al., 2012; Skibniewski, 2014; Cheng and Teizer, 2014; Khoury et al., 2015), computer vision techniques have proven to be efficient in rapidly and

Received September 19, 2017; accepted March 1, 2018

Bahaa Eddine MNEYMNEH, Mohamad ABBAS, Hiam KHOURY
(✉)

Department of Civil and Environmental Engineering, American University of Beirut, Beirut 1107 2020, Lebanon
E-mail: hiam.khoury@aub.edu.lb

conveniently retrieving relevant data from construction sites such as the detection and tracking of workers, material, and equipment (Chi et al., 2009; Park et al., 2012; Memarzadeh et al., 2013; Dimitrov and Golparvar-Fard, 2014; Hamledari et al., 2017). More specifically, recent research efforts have used computer vision and image recognition techniques for construction safety and health monitoring (Seo et al., 2015). For example, Du et al. (2011) introduced the idea of using computer vision techniques to detect hardhats in a video sequence. Their algorithm was divided into two main steps. First, a human face is detected using existing face detection algorithms based on Haar-like features. The system then detects the presence of a hardhat using color segmentation. Their work is considered to be among the first attempts in this field. However, the proposed method was only tested against frontal close-up videos of human faces and did not consider real-case scenarios from construction sites. Similarly, Shrestha et al. (2015) proposed an algorithm that detects workers using standard face detection and then applies edge detection on the region directly above the worker's head. In this case, the system detected a hardhat if its outline was determined to be a semicircle and its color identified as red. However, their system required a set of high-resolution CCTV cameras to be installed on site and was only able to detect hardhats when applied on images captured from the front. Moreover, their system was not assessed on an actual construction site. A more recent effort, Park et al. (2015), detected hardhats using a support vector machine (SVM) classifier as part of a complete framework aimed at enhancing on-site safety conditions. Their algorithm is based on shape recognition and utilizes histogram of oriented gradient (HOG) features to describe the cap-style shape of the hardhat. Although the proposed framework was capable of detecting a hardhat under different conditions and independently from its color, it was also susceptible to false detections because the semi-circular shape of the hardhat could easily be extracted from other irrelevant objects. Many other research studies have targeted similar applications (Gualdi et al., 2011; Bajracharya, 2013; Rubaiyat et al., 2016). However, the aforementioned existing systems were either only able to detect frontal views of hardhats under laboratory conditions and never tested under proper site conditions, were victims of over-prediction and false identification of unwanted objects as hardhats, or were never tested against increasing levels of challenge due to variations in orientation, color, background contrast, image resolution, and on-site lighting conditions. Furthermore, none of the previous studies was concerned with the time efficiency of the detection method for adoption in real-time scenarios. In actuality, the computational speed of the algorithm is as important as its accuracy, especially in safety applications.

Deep learning detection techniques, in particular, neural networks, have gained increased popularity and have been

adopted to analyze digital images in different applications. More specifically, Convolutional Neural Network (CNN) algorithms have demonstrated excellent potential in object detection applications (Krizhevsky et al., 2012). Further research has also been performed to improve the accuracy and reduce the computational time required for these emerging techniques leading to Faster R-CNN (He et al., 2014; Girshick, 2015; Ren et al., 2015). In addition to CNN, other deep learning object detection techniques including YOLO (You Only Look Once), a faster algorithm with reduced precision, and SDD (Single Shot Multibox Detector), have been applied for face, pedestrian, and vehicle detection (Kim et al., 2016; Peng et al., 2016; Zhou et al., 2016). Recently, Fang et al. (2018) applied the Faster R-CNN algorithm to identify non-hardhat use on construction sites. Several experiments were conducted under different conditions and results proved promising whereby high values of precision and recall were achieved. It is important to note that this method requires a large training set of more than 80000 annotated images gathered over more than one year and accordingly necessitates a significant training time.

Consequently, in an attempt to further automate part of the indoor construction safety inspection process and address the aforementioned limitations, this paper aims to evaluate existing manageable object detection techniques for rapidly and efficiently detecting the wearing of hardhats. This evaluation uses standard resolution images captured on actual indoor construction sites under different situations including variations in orientations, colors, background contrast, image resolution, and lighting conditions.

2 Methodology

This section describes and evaluates existing computer vision techniques deemed useful for detecting hardhats. Among these techniques, object detection/recognition methods (Cyganek, 2013) proved promising, in particular: (1) feature detection, extraction, and matching, (2) template-matching, and (3) cascade classifiers models. The usefulness of each visual object recognition method varies according to different factors including but not limited to color, orientation, shape, and scale of the target object. Furthermore, color-based segmentation is explored. The components of the algorithms are implemented using MATLAB 2016a.

2.1 Object detection methods

Local features and descriptors are considered the cornerstone of a large array of computer vision techniques and applications including object detection, tracking, and motion estimation (Dickscheid et al., 2011). More

specifically, in the feature detection, extraction, and matching methods, feature detectors or gradient-based features such as Speeded-Up Robust Features (SURF) or binary features including Binary Robust Invariant Scalable Keypoints (BRISK) and Features from Accelerated Segment Test (FAST) are used to identify point correspondences between the input image and a reference image containing a hardhat (Mikolajczyk and Tuytelaars, 2009). For example, SURF features are first detected from a grayscale image then feature extraction locates the detected features within each image and feature matching identifies similarities between the reference and input images, as indicated in the following code snippet:

```
Surf_Features = detectSURFFeatures(rgb2gray(Image));
[feats1, validpts1] = extractFeatures(rgb2gray(Reference),
    Surf_Features_Reference);
[feats2, validpts2] = extractFeatures(rgb2gray(Input),
    Surf_Features_Input);
Index_Matched_Features = matchFeatures(feats1, feats2).
```

Outliers are then removed and a transformation matrix is calculated using the Random Sample Consensus (RANSAC) algorithm (Oueiss et al., 2012; Khoury et al., 2015). The hardhat having the best match with the reference image is then detected. The number of hardhats is then computed by hiding those identified from the target image such that the next best-matched hardhat can be detected in the next iteration of the algorithm. This counting iterative process halts when no more hardhats can be detected in the target image. It is worth noting that this method functions best for objects displaying non-repeating texture patterns to allow unique and numerous feature matches.

Conversely, template matching frequently refers to a series of operations aimed at detecting and identifying a certain form or pattern in an input image by comparison to a reference or template (Brunelli, 2009). The template is positioned over the input image at every possible position and a similarity coefficient is calculated. Possible metrics to determine the similarity include the sum of absolute differences (SAD), sum of squared differences (SSD), and maximum absolute difference (MaxAD) (Yu et al., 2006). Other methods searching for the minimum difference between two images consist of either an Exhaustive Search (ES) or a Three-step Search (TSS). The former is more accurate, however, more computationally expensive; the latter is faster, yet may not always determine the optimal solution. In MATLAB, a template matcher is typically based on SAD unless otherwise stated (e.g., Three-Step) as indicated in the code snippet below:

```
Detector = vision.TemplateMatcher
    ('SearchMethod','Three-step').
```

A hardhat is detected when the difference computed between a template image containing a hardhat and the input image is less than a required threshold. In general, template-matching algorithms are limited by the available computation power owing to a required high detection accuracy that necessitates lengthy iteration processes.

In this study, different cascade classifiers (Alionte and Lazar, 2015) based on Histogram of Oriented Gradients (HOG), Haar-like, and Local Binary Pattern (LBP) features are assessed. This requires a training process using two sets of positive and negative instances. Positive instances contain images of the relevant object whereas negative instances are images that do not contain the relevant object. A sample of 75 positive and 164 negative images was collected from construction environments to train the three cascade object detectors. The training process also requires a set of input parameters including the number of cascade stages, true positive rate, and false alarm rate (FAR). Experimenting with these parameters yields different results, allowing for the creation of a more effective detector. For example, training a cascade object detector based on HOG features and with the required parameters is performed using the following MATLAB code:

```
trainCascadeObjectDetector('Hog_7_10.XML', positiveInstances, negativeFolder, 'FalseAlarmRate', 0.10, 'NumCascadeStages', 7, 'FeatureType', 'HOG').
```

2.2 Color-based segmentation

Color-based segmentation (Kaur et al., 2013) consists of eliminating positive detections not conforming with possible color schemes of a certain object, in this case a hardhat. To identify those schemes, numerous images of blue, orange, and white hardhats were collected under different lighting conditions and cropped such that the resulting image contained only portions of the hardhat. The average and standard deviation of pixel values were then calculated for each image using RGB, CIE LABORATORY, and HSV color spaces (Kaur et al., 2013) to determine the most accurate representation of the hardhat color.

The RGB color space is defined by the three chromaticities of the red, green, and blue and can produce any chromaticity that is in the triangle defined by these primary colors. The CIE LABORATORY is a color space where L is brightness and a and b are chrominance components, with the difference that the color values are considerably greater than the human gamut; its gamut exceeds that of the RGB. HSV stands for Hue, Saturation, and Value and is based on how colors are conceived by the human vision. Hue refers to pure color form, Saturation refers to the amount of color, and Value refers to the brightness of the color.

3 Object detection methods: preliminary results and analysis

3.1 Performance of feature detection, extraction, and matching algorithm

Owing to the uniform shape and color of a hardhat (e.g., blue or white), the number of detected features was determined to be low (see Fig. 1). One suggested solution to the problem was to add a customized sticker to the hardhat (see Fig. 2a). This then significantly increased the number of extracted features (see Fig. 2b, 2c, 2d).

To further assess the applicability of this algorithm in detecting hardhats in indoor construction environments, experiments were conducted on close-up images (see Fig. 3) clearly indicating the customized stickers on both hardhats. The algorithm is independent from any type of feature used; however, given that a minimum number of features must be extracted with the least computational power, the choice was SURF features.

For example, in Fig. 3a, 63 matching features were identified between the reference and target image in the first iteration of the algorithm. The first detected hardhat was then hidden from the target image and in the second

iteration, 44 matching features were identified between the reference and new target image (see Fig. 3b). The iteration process halts once the second hardhat is hidden and the program returns the final number of detected hardhats. As such, owing to the lack of pertinent features on a hardhat, the algorithm searches for the customized sticker and identifies its target irrespective of the color or shape. However, further testing revealed deficiencies in the system. The method is actually susceptible to misclassifying any object displaying the sticker. Moreover, in a three-dimensional dynamic construction environment, a clear view of the sticker cannot always be guaranteed. In fact, in another sets of experiments, a hardhat could not be detected because either the size or resolution of the sticker was low (see Fig. 4), or the sticker was not visible owing to the orientation of the hardhat. Moreover, the feature extraction and filtering combined with the iteration processes required a relatively high calculation cost.

3.2 Performance of template-matching algorithm

Experiments conducted in an indoor construction environment demonstrated that the algorithm wrongly predicted the location of a blue hardhat when using a template with a

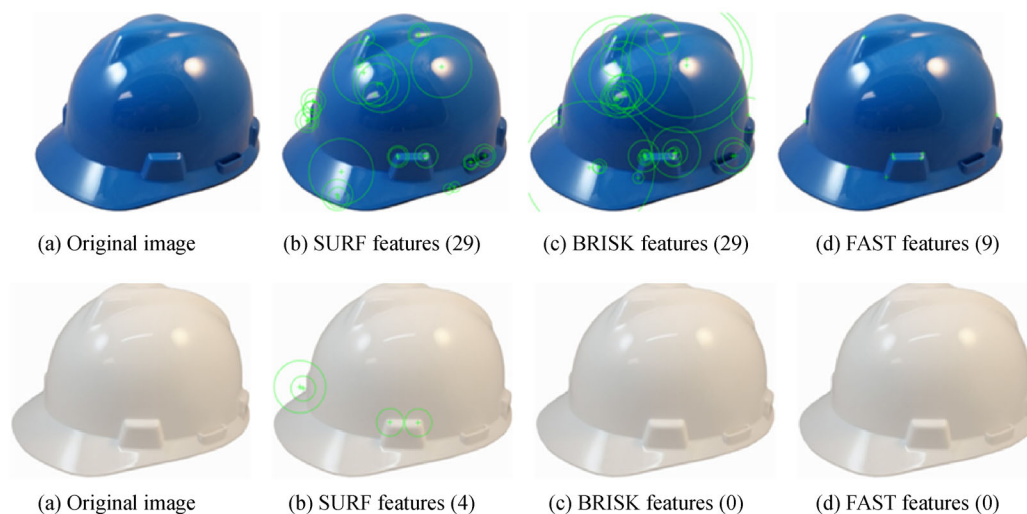


Fig. 1 Number of detected SURF, BRISK, and FAST features in blue (top) and white (bottom) hardhats

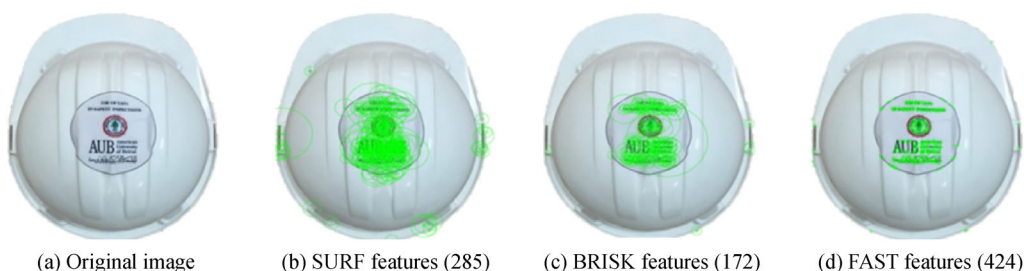


Fig. 2 Number of detected SURF, BRISK, and FAST features in white hardhat with sticker

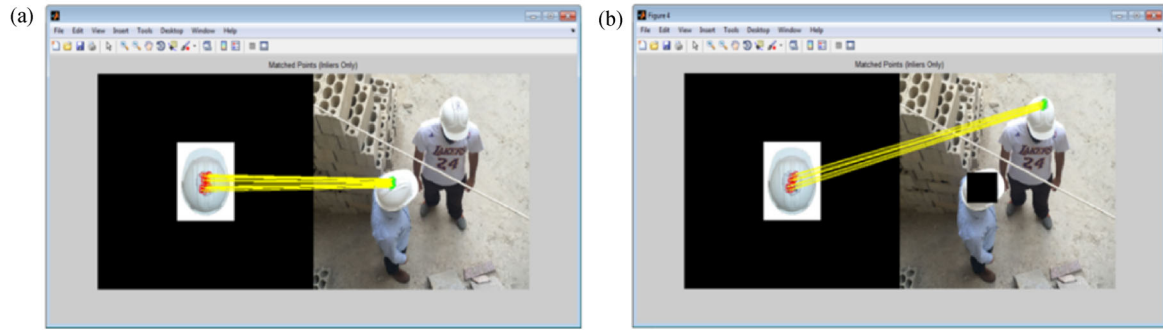


Fig. 3 Matching features for (a) first hardhat and (b) second hardhat

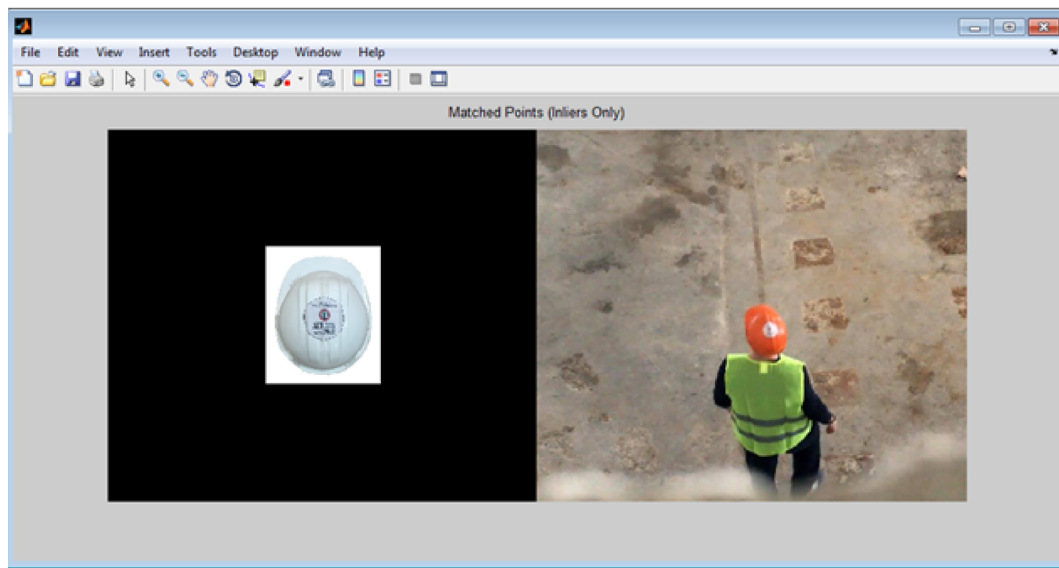


Fig. 4 No detection example—sticker size excessively low

slightly different rotation (see Fig. 5). Hence, a unified template is not sufficient to detect all instances and a classic template matching is relatively inaccurate when addressing any form of difference in scale and rotation. Furthermore, the lengthy calculation process of classic template matching eliminates any usefulness of such an algorithm in a real-time application. In fact, scanning full resolution images from construction sites required hours of processing and significant computational power.

3.3 Performance of cascade classifier

Object detectors are frequently sensitive to out-of-plane transformation. However, this should not be a problem in the case of hardhat detection because its semi-circular shape remains unchanged regardless of the viewing angle. Cascade detectors based on Haar and LBP features yielded high rates of wrong detection in all testing images (see Fig. 6). Conversely, detectors based on HOG features could accurately describe the circular shape of the hardhat

irrespective of its color (see Fig. 7).

The ability of the detector to correctly identify hardhats from different viewpoints was further verified using a set of three testing images containing front, side, and back views of a blue hardhat (see Fig. 8). The classifier was also capable of recognizing two objects simultaneously. Color variations also had no effect and the computational speed was acceptable. Furthermore, the capacity to experiment with training parameters to obtain different results is another main advantage of cascade classifiers.

3.4 Discussion, comparison, and selection

Based on preliminary experiments and results featured above, the three computer vision techniques were assessed. The first evaluated technique, feature detection, extraction, and matching, is useful for identifying specific common features between a reference and input image. However, a considerably limited number of features was extracted from a hardhat due to its uniform shape. Hence, a



Fig. 5 Wrong detection using template matching

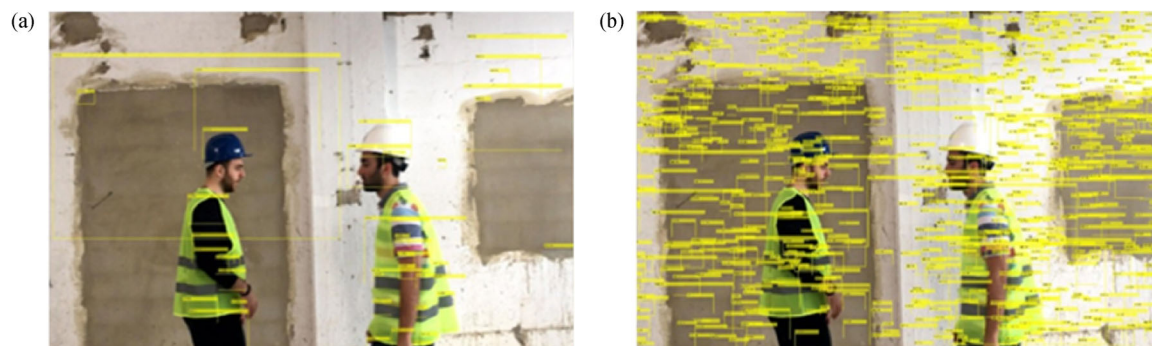


Fig. 6 High rate of incorrect detections using (a) Haar features and (b) LBP features

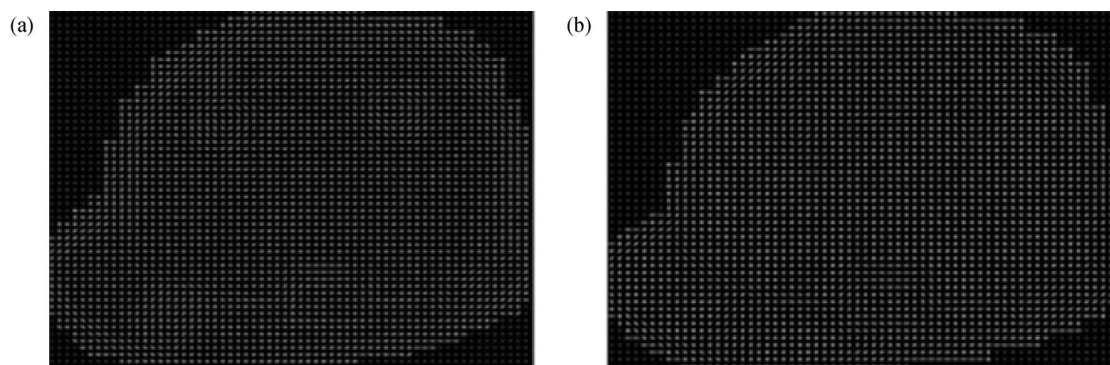


Fig. 7 Extracted HOG features: (a) blue hardhat and (b) white hardhat

customized sticker was added to improve the results. Nevertheless, false detections occurred whenever the sticker was not clearly visible, or its size or resolution

was low. In the second evaluated technique, template matching, an object is identified by calculating the similarity coefficient between a testing and target image.

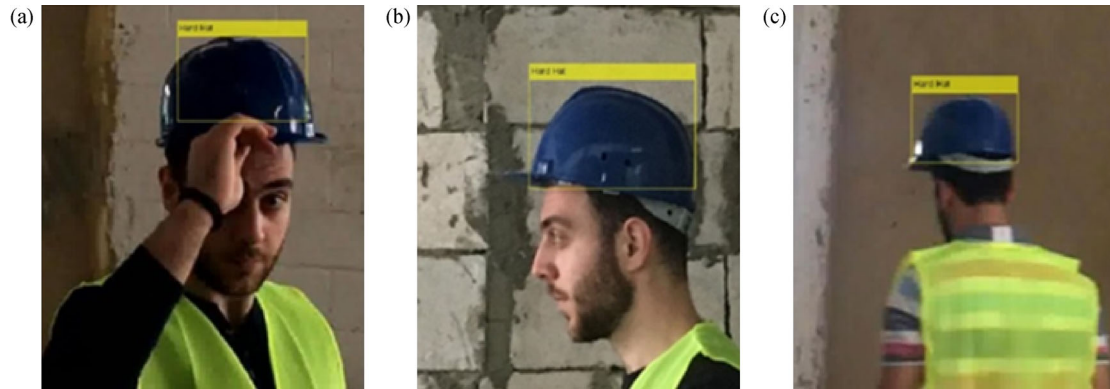


Fig. 8 Detected hardhat: (a) front view, (b) side view, and (c) back view

However, the inability to detect hardhats from different orientation angles made this technique impractical for a real-time application in an actual dynamic environment. Conversely, cascade classifiers are trained using a set of positive and negative images, then used to detect objects based on specific features. This third technique proved robust against changes in orientation, size, and color. These findings are comparatively summarized in Table 1. Accordingly, the cascade object detector, in particular the HOG-based detector, clearly outperformed the other presented object detection techniques and can be potentially adopted in real-time safety applications.

4 Color-based segmentation methods: preliminary results and analysis

For a sample of 102 blue hardhats, the average values of Red, Green, and Blue in the RGB representation were relatively low, describing the “dark” rather than the “blue” color of the hardhat. Red, Green, and Blue values spanning from 2.5 to 36, 3.9 to 57, and 21 to 100, respectively (see Fig. 9), do not well represent the blue color. Similar insignificant results were obtained using the CIE LABORATORY color space. For the HSV color space, the calculated averages of Hue, Saturation, and Value for the 102 sample images ranged from 0.59 to 0.66, 0.66 to 0.95, and 0.08 to 0.39, respectively (see Fig. 10), while providing the most accurate representation of a blue color with low brightness. The values of standard deviation for

Hue were also minimal (see Fig. 11), implying that the uniformity of the hardhat color within each image was accurately modeled. The same procedure was applied for other colors of hardhats (e.g., orange and white) and the best results were obtained using the HSV color space.

Hence, in this study, color-based segmentation in HSV color space is adopted. In addition, for best results, it was decided to apply this technique on the parts of the images identified as hardhats from the HOG-based cascade classifier process with the aim of reducing false detections. This potential combination of algorithms thereby warrants further experimentation.

5 Experimental analysis of cascade classifier

In this section, further assessment of the HOG-based cascade object detector was performed. Two seven-stage cascade object detectors were trained using the same image data sets of 75 positive and 164 negative images and two different values for the FAR, 0.05 and 0.1. In theory, a greater FAR should yield more false positive results and the detector should be less likely to miss a desired object. The performance of the cascade classifier was then analyzed against variations in orientation and color, background contrast, image resolution, and lighting conditions. Three scenarios were devised and the respective results are depicted in Tables 2, 3, and 4 including computational durations (t) together with their underlying

Table 1 Comparative summary of object detection techniques

| Computational duration | Feature detection, extraction, & matching | Medium | Template matching | Very high | Cascade classifier | Low |
|----------------------------------|---|--------|-------------------|-----------|--------------------|-----|
| Color invariance | Yes | No | No | Yes | | |
| Orientation invariance | No | No | No | Yes | | |
| Practicality | No | No | No | Yes | | |
| Customizability | Yes | No | No | Yes | | |
| Training database/fingerprinting | No | No | No | Yes | | |

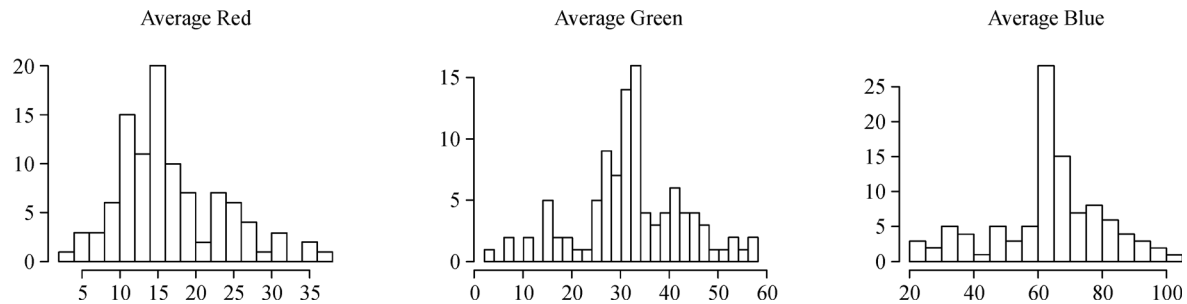


Fig. 9 Average Red, Green, and Blue values in images of blue hardhat

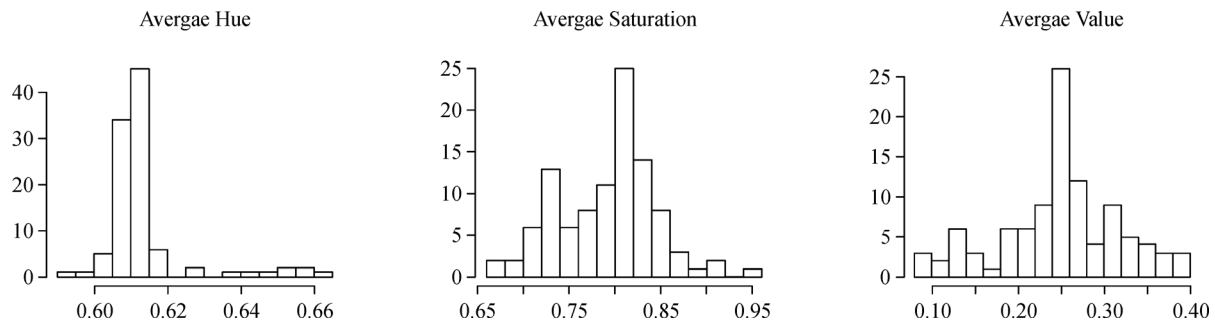


Fig. 10 Average Hue, Saturation, and Value values in images of blue hardhat

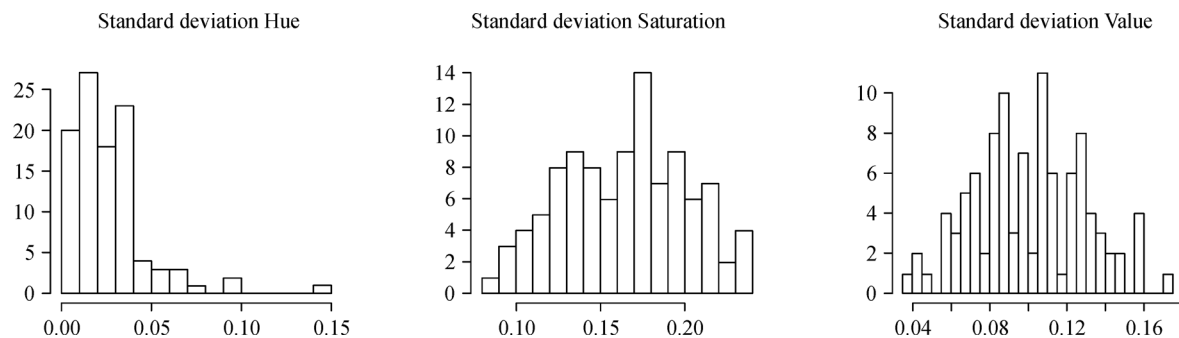


Fig. 11 Standard deviation of Hue, Saturation, and Value values in images of blue hardhat

time statistics.

In Scenario 1, the level of challenge was relatively low and all hardhats could be easily discerned from their respective backgrounds. The two detectors were tested on ten images with 13 hardhats in total; the one with the FAR set to 0.05 missed three hardhats, whereas the detector with the FAR set to 0.1 did not miss any hardhats (Table 2). Nevertheless, both classifiers were subject to wrong identification and for example, in Image 6, a mobile worker's head was mistakenly classified as a hardhat (see Fig. 12). The computational durations of both detectors were similar with an average processing time per image of approximately 2 s.

For Scenario 2, the level of challenge was significantly increased. The low contrast between the hardhat and its

background (e.g., white hardhat in front of a white wall) could reduce the significance of the detected HOG features, which, in turn, could reduce the efficiency of the detector. In fact, the performance of the cascade object detector did reduce compared to Scenario 1. More specifically, for the higher FAR of 0.1, the detector could identify ten out of 13 hardhats (Table 3). That is, the detector did remain efficient even when the contrast with the background was minimal (see Fig. 13). The computational durations of both detectors did not significantly vary with a similar average processing time per image of approximately 2 s.

To assess the effect of changing the resolution of the test images on the detection results independently of other factors, the third experiment (Scenario 3) was performed

Table 2 Performance of cascade classifier in Scenario 1

| Scenario 1–High contrast against background, variable colors, and orientations | | | | | | | | | | | |
|--|-----------------|---|--------------|---|--------------------------|---|---|----------------|---|----|--------------------|
| Image ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| True number of hardhats | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | |
| Detected-FAR = 0.05 | 1 | 1 | 0 | 0 | 1 | 2 | 1 | 2 | 1 | 3 | <i>t</i> = 2.171 s |
| Detected-FAR = 0.1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 3 | <i>t</i> = 2.231 s |
| <i>Time statistics-FAR = 0.05</i> | | | | | | | | | | | |
| Action | Number of Calls | | Total Time/s | | Percentage of Total Time | | | Average Time/s | | | |
| Execution of detector | 10 | | 19.949 | | 91.9% | | | 1.9949 | | | |
| Reading image file | 10 | | 1.734 | | 8.0% | | | 0.1734 | | | |
| All other actions | - | | 0.031 | | 0.1% | | | 0.0031 | | | |
| Total | - | | 21.714 | | 100% | | | 2.1714 | | | |
| <i>Time statistics-FAR = 0.1</i> | | | | | | | | | | | |
| Action | Number of Calls | | Total Time/s | | Percentage of Total Time | | | Average Time/s | | | |
| Execution of detector | 10 | | 20.517 | | 91.9% | | | 2.0517 | | | |
| Reading image file | 10 | | 1.767 | | 7.9% | | | 0.1767 | | | |
| All other actions | - | | 0.034 | | 0.2% | | | 0.0034 | | | |
| Total | - | | 22.318 | | 100% | | | 2.2318 | | | |

Table 3 Performance of cascade classifier in Scenario 2

| Scenario 2- Low contrast against background, variable colors, and orientations | | | | | | | | | | |
|--|-----------------|---|--------------|---|--------------------------|---|---|----------------|---|-----------------------|
| Image ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| True number of hardhats | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 |
| Detected- FAR = 0.05 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | $t = 2.191 \text{ s}$ |
| Detected- FAR = 0.1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 2 | 2 | $t = 2.196 \text{ s}$ |
| <i>Time statistics-FAR = 0.05</i> | | | | | | | | | | |
| | | | | | | | | | | |
| Action | Number of Calls | | Total Time/s | | Percentage of Total Time | | | Average Time/s | | |
| Execution of detector | 10 | | 20.120 s | | 91.8% | | | 2.0120 s | | |
| Reading image file | 10 | | 1.765 s | | 8.1% | | | 0.1765 s | | |
| All other actions | - | | 0.030 s | | 0.1% | | | 0.0030 s | | |
| Total | - | | 21.915 s | | 100% | | | 2.1915 s | | |
| <i>Time statistics-FAR = 0.1</i> | | | | | | | | | | |
| | | | | | | | | | | |
| Action | Number of Calls | | Total Time/s | | Percentage of Total Time | | | Average Time/s | | |
| Execution of detector | 10 | | 20.148 s | | 91.7% | | | 2.0148 s | | |
| Reading image file | 10 | | 1.778 s | | 8.1% | | | 0.1778 s | | |
| All other actions | - | | 0.036 s | | 0.2% | | | 0.0036 s | | |
| Total | - | | 21.962 s | | 100% | | | 2.1962 s | | |

using the same images of Scenario 1, however, cropped or resized to obtain an image resolution of (1920×1080) pixels. Images 1 to 5 were cropped; Images 6 to 10 were resized. Cropped images (1–5) yielded results identical to Scenario 1. Conversely, resizing the image could decrease the size of the hardhats below the trained size and accordingly prevent hardhat detection. This was the case in Images 7 and 8 (Table 4). Therefore, training the

detector using images of the same resolution and captured from a similar distance as the test images can positively influence detection results. Compared to Scenarios 1 and 2, these scenario's computational durations improved significantly from approximately 2 s to approximately 0.5 s.

A fourth scenario (i.e., low luminosity) was also considered. However, its results were not reported because

Table 4 Performance of cascade classifier in Scenario 3

| Scenario 3–Different image resolutions | | | | | | | | | | | |
|--|-----------------|---|--------------|---|--------------------------|---|---|----------------|---|----|--------------------|
| Image ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
| True number of hardhats | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | |
| Detected- FAR = 0.05 | 1 | 1 | 0 | 0 | 1 | 2 | 0 | 1 | 1 | 3 | <i>t</i> = 0.541 s |
| Detected- FAR = 0.1 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 1 | 2 | 3 | <i>t</i> = 0.568 s |
| <i>Time statistics- FAR = 0.05</i> | | | | | | | | | | | |
| Action | Number of Calls | | Total Time/s | | Percentage of Total Time | | | Average Time/s | | | |
| Execution of detector | 10 | | 4.887 | | 90.3% | | | 0.4887 | | | |
| Reading image file | 10 | | 0.486 | | 9.0% | | | 0.0486 | | | |
| All other actions | - | | 0.040 | | 0.7% | | | 0.0040 | | | |
| Total | - | | 5.413 | | 100% | | | 0.5413 | | | |
| <i>Time statistics- FAR = 0.1</i> | | | | | | | | | | | |
| Action | Number of Calls | | Total Time/s | | Percentage of Total Time | | | Average Time/s | | | |
| Execution of detector | 10 | | 5.150 | | 90.7% | | | 0.5150 | | | |
| | 10 | | | | 8.6% | | | | | | |
| All other actions | - | | 0.040 | | 0.7% | | | 0.0040 | | | |
| Total | - | | 5.681 | | 100% | | | 0.5681 | | | |

**Fig. 12** Wrong classification of head region in Image 6–Scenario 1

the variation in image luminosity did not considerably influence the performance of the HOG-based detector, as it is capable of describing the shape of the object irrespective of its color.

6 Experimental analysis of the color-based segmentation method

To test the effectiveness of the proposed algorithm, the

developed tool was applied on images where the cascade classifier incorrectly identified unwanted objects as hardhats. The results highlighted the capability of the color-based segmentation tool to eliminate false detections by filtering out the objects not matching with any of the aforementioned color statistics. For different runs, the algorithm accurately detected and identified the hardhat object in the images and eliminated any previous false detections output from the HOG-based cascade classifier (see Fig. 14).



Fig. 13 Correct identification of both hardhats in Image 8–Scenario 2



Fig. 14 Hardhat detection results for positive images: (a) cascade object detector and (b) cascade object detector combined with color-based segmentation

7 Conclusions, limitations, and future work

This paper evaluated existing computer vision algorithms, in particular object detection and color-based segmentation methods, in efficiently and rapidly detecting hardhats on indoor construction sites. Several experiments were conducted and the results revealed that a well-trained cascade classifier was found to be robust under different scenarios and conditions. Furthermore, it was proven to be relatively time efficient and in a real-time application, it would be capable of scanning for violations every two seconds. The process can actually be expedited by reducing the resolution of the training and test images. Moreover, color-based segmentation proved effective in reducing false detections output from the cascade classifier process.

While this research study achieved promising results under different scenarios and situations, it exhibited limitations. The performance evaluation of the proposed algorithms was limited to a relatively small number of testing images. Furthermore, all experiments were performed using a single type of hardhat and three different colors. Moreover, the study did not consider the cases where multiple hardhats could be detected in a single region of interest or hardhats could be occluded by others. As such, a more comprehensive testing set is required and the robustness of the computer vision algorithms against other variations in the shape and color of the hardhat along with occlusions must be investigated. Further work will investigate improving the accuracy and recall of the hardhat detection process and eliminating false detections by combining a weak cascade classifier with other image and color-based segmentation techniques. Further testing is also required to evaluate the accuracy of other object detection algorithms and to explore the potential use of heat cameras in addition to digital imagery. Future studies will aim as well at integrating this hardhat detection process into a complete safety inspection framework capable of rapidly, efficiently, and conveniently issuing safety warnings when a safety violation is detected and alerting nearby workers of hazards.

Acknowledgements The presented work was supported by AUB's University Research Board (URB). The authors gratefully acknowledge URB support. Any opinions, findings, conclusions, and recommendations expressed by the authors in this paper do not necessarily reflect the views of URB.

References

Abbas M, Mneymneh B E, Khoury H (2016). Use of unmanned aerial vehicles and computer vision in construction safety inspections. In: Proceedings of the 3rd Australasia and South-East Asia Structural Engineering and Construction Conference (ASEA-SEC-3). Kuching, Sarawak, Malaysia

- Abbas M, Mneymneh B E, Khoury H (2018). Assessing on-site construction personnel hazard perception in a middle eastern developing country: An interactive graphical approach. *Safety Science Journal*, Elsevier Science, 103: 183–196
- Alionte E, Lazar C (2015). A practical implementation of face detection by using matlab cascade object detector. In: Proceedings of Conference on System Theory, Control and Computing. Cheile Gradistei, Romania, 785–790
- Bajracharya D (2013). Real time pattern recognition in digital video with application to safety in construction sites. Dissertation for the College Degree. Las Vegas: University of Nevada–Las Vegas
- Brunelli R (2009). *Template Matching Techniques in Computer Vision: Theory and Practice*. New Jersey: Wiley
- Chdid D, Oueis R, Khoury H, Asmar D, Elhajj I (2011). Inertial-vision sensor fusion for pedestrian localization. In: Proceedings of IEEE International Conference on Robotics and Biomimetics (ROBIO). 1695–1701
- Cheng T, Teizer J (2014). Modeling tower crane operator visibility to minimize the risk of limited situational awareness. *Journal of Computing in Civil Engineering*, 28(3): 04014004
- Chi S, Caldas C H, Kim D Y (2009). A methodology for object identification and tracking in construction based on spatial modeling and image matching techniques. *Computer-Aided Civil and Infrastructure Engineering*, 24(3): 199–211
- Cyganek B (2013). *Object Detection and Recognition in Digital Images: Theory and Practice*. New Jersey: Wiley
- Dickscheid T, Schindler F, Förstner W (2011). Coding images with local features. *International Journal of Computer Vision*, 94(2): 154–174
- Dimitrov A, Golparvar-Fard M (2014). Vision-based material recognition for automated monitoring of construction progress and generating building information modeling from unordered site image collections. *Advanced Engineering Informatics*, Elsevier, 28 (1): 37–49
- Ding L Y, Yu H L, Li H, Zhou C, Wu X G, Yu M H (2012). Safety risk identification system for metro construction on the basis of construction drawings. *Automation in Construction*, Elsevier, 27: 120–137
- Du S, Shehata M, Badawy W (2011). Hardhat detection in video sequences based on face features, motion and color information. In: Proceedings of 3rd IEEE International Conference on Computer Research and Development
- Fang Q, Li H, Luo X, Ding L, Luo H, Rose T M, An W (2018). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction*, 85: 1–9
- Gheisari M, Irizarry J, Walker B N (2014). UAS4SAFETY: The potential of unmanned aerial systems for construction safety applications. *Construction Research Congress*
- Girshick R (2015). Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. 1440–1448
- Gualdi G, Prati A, Cucchiara R (2011). Contextual Information and covariance descriptors for people surveillance: An application for safety of construction workers. *EURASIP Journal on Image and Video Processing*, 2011(1): 1–16
- Ham Y, Han K K, Lin J J, Golparvar-Fard M (2016). Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): A review of related works. *Visualization in*

- Engineering, 4(1): 1
- Hamledari H, McCabe B, Davari S (2017). Automated computer vision-based detection of components of under construction indoor partition. *Automation in Construction*, Elsevier, 74: 78–94
- He K, Zhang X, Ren S, Sun J (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. *European Conference on Computer Vision*, 346–361
- Health and Safety Executive (2014). Health and safety in construction sector in Great Britain. <http://www.hse.gov.uk/statistics/industry/construction/construction.pdf>
- Kaur N, Banga V K, Kaur A (2013). Image segmentation based on color. In: *Proceedings of International Journal of Research in Engineering and Technology*. 2(11)
- Khoury H, Chdid D, Oueis R, Elhadj I, Asmar D (2015). Infrastructure-less approach for ubiquitous user location tracking in construction environments. *Automation in Construction*, 56: 47–66
- Khoury H M, Akula M, Kamat V R (2012). *Mobile and Pervasive Computing in Construction*. Hoboken: John Wiley & Sons
- Khoury H M, Kamat V R (2009a). Evaluation of position tracking technologies for user localization in indoor construction environments. *Automation in Construction*, Elsevier Science, 18(4): 444–457
- Khoury H M, Kamat V R (2009b). High-precision identification of contextual information in location-aware engineering applications. *Advanced Engineering Informatics*, Elsevier Science, 23(4): 483–496
- Kim H, Lee Y, Yim B, Park E, Kim H (2016). On-road object detection using deep neural network. In: *Proceedings of IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*. IEEE, 1–4
- Kim Y S, Lee J H, Yoo H S, Lee J B, Jung U S (2009). A performance evaluation of a Stewart platform based Hume concrete pipe manipulator. *Automation in Construction*, 18(5): 665–676
- Kopp A P, Saidi K S, Khoury H M (2010). Evaluation of ultra-wideband technology for use in 3D locating systems. In: *Proceedings of the 10th Performance Metrics for Intelligent Systems (PerMIS) Workshop*. Baltimore, MD, United States, 7–12
- Krizhevsky A, Sutskever I, Hinton G E (2012). Imagenet classification with deep convolutional neural networks. In: *Proceedings of the International Conference on Neural Information Processing Systems*. 1097–1105
- Memarzadeh M, Golparvar-Fard M, Niebles J C (2013). Automated 2D detection of construction equipment and workers from site video streams using histograms of oriented gradients and colors. *Automation in Construction*, 32: 24–37
- Mikolajczyk K, Tuytelaars T (2009). *Encyclopedia of Biometrics*. New York: Springer US
- Mneymneh B E, Abbas M, Khoury H (2016). A UAV-based image processing system for identifying and visualizing construction hazardous areas. In: *Proceedings of 16th International Conference on Construction Applications of Virtual Reality*. Hong Kong Occupational Safety and Health Administration
- Commonly Used Statistics. <https://www.osha.gov/oshstats/common-stats.html>, 2016–12–19
- Oueiss R, Chdid D, Khoury H M, Asmar D, Elhadj I (2012). Infrastructureless inspection process automation in construction. *International Society for Gerontechnology*, 11(2): 92–98
- Park M W, Elsafty N, Zhu Z (2015). Hardhat-wearing detection for enhancing on-site safety of construction workers. *Journal of Construction Engineering and Management*, 141(9)
- Park M W, Koch C, Brilakis I (2012). Three-dimensional tracking of construction resources using an on-site camera system. *Journal of Computing in Civil Engineering*, 26(4): 541–549
- Peng Q, Luo W, Hong G, Feng M, Xia Y, Yu L, Hao X, Wang X, Li M (2016). Pedestrian detection for transformer substation based on gaussian mixture model and YOLO. In: *Proceedings of International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*. IEEE, 562–565
- Ren S, He K, Girshick R, Sun J (2015). Faster r-cnn: towards real-time object detection with region proposal networks. In: *Proceedings of International Conference on Neural Information Processing Systems*. 91–99
- Rubaiyat A H M, Toma T T, Kalantari-Khandani M, Rahman S A, Chen L, Ye Y, Pan C S (2016). Automatic detection of helmet uses for construction safety. In: *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence Workshops*. 135–142
- Seo J, Han S, Lee S, Kim H (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, Elsevier, 29(2): 239–251
- Shrestha K, Shrestha P P, Bajracharya D, Yfantis E A (2015). Hard-hat detection for construction safety visualization. *Journal of Construction Engineering*, 1–8
- Skibniewski M J (2014). Information technology applications in construction safety assurance. *Journal of Civil Engineering and Management*, 20(6): 778–794
- Chae S, Yoshida T, (2010). Application of RFID technology to prevention of collision accident with heavy equipment. *Automation in Construction*, 19(3): 368–374
- Yu J, Amores J, Sebe N, Tian Q (2006). A new study on distance metrics as similarity measurement. In: *Proceedings of IEEE International Conference on Multimedia and Expo*
- Zhou Y, Nejati H, Do T T, Cheung N M, Cheah L (2016). Image-based vehicle analysis using deep neural network: A systematic study. In: *Proceedings of IEEE International Conference on Digital Signal Processing*