Application of vision systems to improve the effectiveness of monitoring compliance with technical safety requirements at industrial facilities

A A Ekimenko[1], E A Ekimenko[1] and S V Shavetov[1]

[1]ITMO, Kronverksky prospect, 49, Saint Petersburg, Russia, 197101

e-mail: ekimenkoanton@mail.ru, modest3005@yandex.ru

Abstract. According to the Federal State Statistics Service for 2017, over 25,000 injuries at work were registered in Russia, of which 1,138 were fatal. Studies have shown that most of the injuries in the workplace due to non-compliance with technical safety rules, namely the lack of personal protective equipment or their improper use. To improve the efficiency of monitoring compliance with the rules of technical safety at industrial facilities, this paper discusses the use of vision systems for automatic control of the availability of personal protective equipment at workers in the area of industrial work.

1. Introduction

Construction is a high-risk activity that requires builders to lift heavy construction materials, work in uncomfortable poses, and perform high-intensity operations that are key factors leading to workplace injuries [1]. The consequences of head and neck injuries are the most serious, and often fatal [2]. Wearing a helmet is an effective protection measure to minimize the risk of traumatic brain injury. Helmets protect workers, prevent the penetration of sharp objects, absorb impact from blunt objects and reduce the risk of electric shock. Despite the vital role of helmets in the protection of life, most of the workers who received head injuries at the time of the incident did not have head protection [3].

The method of automated monitoring of personal protective equipment(PPE) on workers can improve safety at the production site. However, the currently existing methods of detecting the absence of helmets on workers have significant limitations, and many cannot be used in real-time monitoring systems. Some existing methods have proven themselves to work in the near field, but they are not very effective in detecting people at a long distance. This is due to the fact that the resolution of the workers in the image is too small to extract facial features that are clearly visible in near-field frames. [4]

Most surveillance cameras of construction sites are installed at the border of the construction site, at high altitude. Long-distance video is distinguished by the low resolution of workers (the area not more than 30 pixels) in the image, a wide background and various poses of people [5,6], which is a serious problem when detecting people without a hard hat on construction sites.

This article discusses the method of recognizing people without helmets in the far field on open production sites.

In order to test the robustness of the method to the changing conditions of the construction site, this study also analyzes various visual factors that have a negative effect on the detection process.

2. Existing PPE control methods

Currently, research into the detection of the absence of a hard hat can be divided into two categories: tag(sensors), based detection methods and computer vision-based detection methods (RFID). In 2013, A. Kelm [7] proposed a mobile radio frequency identification method to verify the compliance of personnel equipment with technical safety requirements. In the proposed method, the monitoring system of personal protective equipment consists of a set of RFID tags associated with the worker's respective PPE, such as hard hats, goggles, respirators, etc., and a user interaction module performing the task of reading RFID tags, which by the presence of the appropriate labels determine what kind of PPE the worker is equipped with. To determine the state of the user's protective equipment, the set of reading items are compared with the previously saved list of PPE that worker must have to be able to enter the facility. Since RFID readers were located at the entrance to the construction site, this system can only guarantee that some worker entered the construction site with the necessary PPE. This system does not allow controlling the use of PPE by a worker after passing a checkpoint at the production site. In addition, marking the PPE with an RFID tag only indicates that distance between the worker and the PPE is close, but does not guarantee that PPE is used properly by the worker. Later, S. Barro-Torres proposed a new monitoring system for PPE that allows for monitoring the use of PPE throughout the production site. Instead of checking PPE at the entrance, it was proposed to equip workers with devices that collect information about the presence of PPE locally and transmit this information to the data aggregation server using a wireless data channel such as WI-FI, Bluetooth or mobile communication. This system also, like the previous one, does not allow determining whether the worker is wearing a hard hat on the head or it is just next to it [9]. A pressure sensor is installed in the hard hat to determine if the hard hat was worn, the collected information is transmitted via Bluetooth wireless communication to the monitoring system server. If the employee goes beyond the permissible range of data transmission from the receiver device for a long period — the data may be lost, since the data storage volume locally on the sensor is limited and has a period of overwriting, which makes it impossible to determine the presence of PPE in some time intervals, information about which did not have time to be transferred to the server. In addition, these devices must be charged regularly. The need for a regular charge can limit its use and damage the widespread use of this technology.

In view of the above, the use of existing methods of detecting and tracking PPE based on tags and sensors is limited by the need for each builder to use tags or sensors. This can be considered as an obsessive requirement for workers, besides the use of this technology requires large financial investments in additional equipment, such as tags, sensors, readers and transmitters. Many workers do not want to wear such equipment because of possible health problems or privacy issues.

Methods for monitoring PPE based on image recognition have become more common, thanks to more extensive monitoring capabilities. RGB-D cameras, such as Kinect and XTION, are one of the most popular tools for analyzing the deviant behavior of employees [10-12]. However, the range of operation of these cameras is strongly limited by the distance from 1 to 4 m [13], which does not allow their use in open areas. Also, sensors of this type are susceptible to interference from sunlight and ferromagnetic radiation, which makes them unsuitable for detecting the absence of hard hats on construction sites [14]. In this regard, the use of conventional cameras, especially single camera, has a competitive advantage for practical use. However, there are still problems with automatic detection systems for the absence of hard hats using one or more cameras. For example, S. Du offers a method for detecting the absence of hard hats, based on facial features, information about movement and color [15]. Color-based face recognition methods have two important assumptions: all workers turn to face the camera while working, and all hard hats have the same color. In practice, these two assumptions may not be fulfilled on a real construction site. K. Shresta suggests using edge detection algorithms to recognize the edges of objects in the upper area of the face, where the hard hat can be recognized [16]. This method also depends on facial recognition and does not produce a positive result if the employee does not face the camera. A. H. Rubayat proposes a method for detecting the absence of hard hats, using histograms of oriented gradients (HOG) and Hough transform to obtain characteristics of workers and hard hats [17,18]. Like the previous ones, this method is also based on facial features detection and has similar limitations. To solve the limitations of the above methods, this article proposes an algorithm for automatic detection of the absence of hard hats, based on the Faster R-CNN. The R-CNN Faster algorithm takes less time to process an image and has higher precision than previous methods.

3. The proposed method of control of personal protective equipment

Open production sites are very complex environments. A variety of weather events, changes in light, changes in the distance to the object, the occurrence of overlaps and changes in people's poses can have a significant impact on the quality of worker detection in a long-distance observation system.

Based on the analysis of existing solutions made in Chapter 2, we can conclude that existing methods of detection based on computer vision are limited in their practical application in real-world scenarios.

Thus, the overall goal of this work is to develop a new method for monitoring the use of workers' PPE on open production sites and assess the possibility of using the proposed method to detect the absence of hard hats in various conditions.

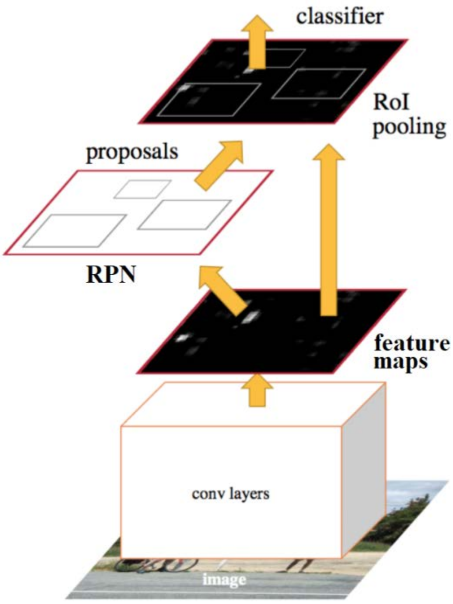

Figure 1. Faster R-CNN algorithm.

Faster R-CNN is a neural network object detection method proposed by S. Ren in 2015 [19]. In Faster R-CNN, the Region Proposal Network (RPN) is introduced instead of the slow selective search algorithm, that can generate high-quality candidate regions that are used to detect and classify objects. Faster R-CNN allows processing up to 5 frames per second, that, in combination with high precision, allows using this method in real object detection systems.

Faster R-CNN consists of three main steps, as shown in Figure 1. The first step to obtain a feature map of the objects, the original image is fed to the input of the CNN network represented by multiple layers of the convolutions and layers under the samples (MaxPooling). The next step is a fully convolutional RPN network, in which the obtained features are used to form candidate regions. Since the whole picture contains many unnecessary objects, and people always appear very small, it is difficult to draw a conclusion based on the feature map, whether a person wears a helmet or not. Therefore, only foreground areas will be used to recognize the absence of hard hats. The third step is represented by the Fast R-CNN classifier, which uses the obtained regions and the corresponding attributes to determine whether the given region contains a person without hard hats. Compared to other methods used to detect the absence of helmets, the Faster R-CNN has three advantages:

- Faster R-CNN is more reliable in working with complex scenes of open industrial sites. In particular, the previously used methods work only if the person is in a standing position. While the Faster R-CNN method is capable of reliably operating in the independence of human pose, workplace position, light, visibility range, and partial overlaps.
- Faster R-CNN has a higher precision that meets the requirements of engineering problems. As a result of testing the method on the Pascal VOC 2012[20] dataset, the precision of 89.6% was obtained, while for the HOG — 10.2 % [21].
- Thirdly, this method has a high speed of operation, which allows creating a system of monitoring the absence of hard hats, working in real time.

Thus, on the basis of the above, we conclude that Faster R-CNN method has higher rates of robustness, precision, and speed compared with the previously used, meeting the practical requirements for safety monitoring at various industrial facilities.

3.1 Performance metrics

The effectiveness of the method was evaluated based on its precision, speed, and robustness as follows. To assess the quality of work, we will use the following metrics: precision — the ability of the method to distinguish this class (workers without hard hats) from other classes, recall – the ability of the algorithm to detect workers without hard hats, miss rate (probability of skipping) – the opposite of recall indicates what percentage of workers without helmets was skipped by the method.

$$\begin{aligned} precision &= \frac{TP}{TP + FP} \\ recall &= \frac{TP}{TP + FN} \\ miss\,rate &= 1 - recall = \frac{FN}{TP + FN} \end{aligned} \tag{1}$$

where TP — number of correctly accepted hypotheses (workers without hard hats), FP — number of incorrectly accepted hypotheses (workers in hard hats), FN — number of incorrectly rejected hypotheses (workers without hard hats are not detected).

The speed of the method will be estimated as the amount of time required to detect the worker without a hard hat for one image.

Robustness represents the degree of stability of the method when working on various images. Open production sites usually contain many workers, equipment and building materials. Also, on open production sites, changes in weather, lighting, visibility often occur, there are partial overlaps of the object of detection. These factors have a significant impact on the visual perception of such scenes and, accordingly, on the results of detection. A good algorithm should be resistant to such changes.

4. Experiments and results

Training and testing of the method were carried out on a data set of 100,000 objects representing images from various open production sites. The training set is made up of 80,000 randomly selected images, testing was conducted on the remaining 20,000 images.

The effectiveness of the method was evaluated using the above-described metrics of quality, speed, and robustness. To calculate these metrics, all images in the test data set were divided into several categories (Table 1) – external factors, impacting the correct detection, such as weather changes, lighting, people's poses, range, and overlap. The next step was calculating the above-mentioned metrics for each category.

The Faster R-CNN model returns a probabilistic value for each detected object. The probability value here is defined as the probability that an object is a worker without hard hats. For example, a probability value of 0.9 means that probability that an object is a worker without a hard hat is 90%. An event is considered positive if the resulting probability exceeds the specified probability threshold. Therefore, the probability threshold value affects the classification of positive and negative events.

As shown in figure 2, a high probability threshold allows excluding ambiguous events from the results, which leads to high precision, but low recall, while low threshold allows more ambiguous events, which gives a high level of recall but low precision. For optimal results, both in precision and recall, a probability threshold of 0.7 was chosen.

Table 1. Information about image datasets for different categories.

Categories	Subcategories	Number of workers without hard hats	Number of images
Weather	Sunny	3134	1000
	Cloudy	2783	1000
	Rainy	1559	1000
	Foggy	2217	1000
Light	8:00 - 10:00	2831	1000
	10:00 - 13:00	3125	1000
	13:00 - 16:00	2952	1000
	16:00 - 18:00	2127	1000
Pose	Standing	2542	1000
	Bending	1762	1000
	Squatting	1275	1000
	Sitting	1128	1000
Distance	Short	2136	1000
	Medium	2854	1000
	Long	2916	1000
Overlap	Whole body visible	1924	1000
	Upper body visible	1015	1000
	Head visible	1393	1000
	Part of head visible	1067	1000

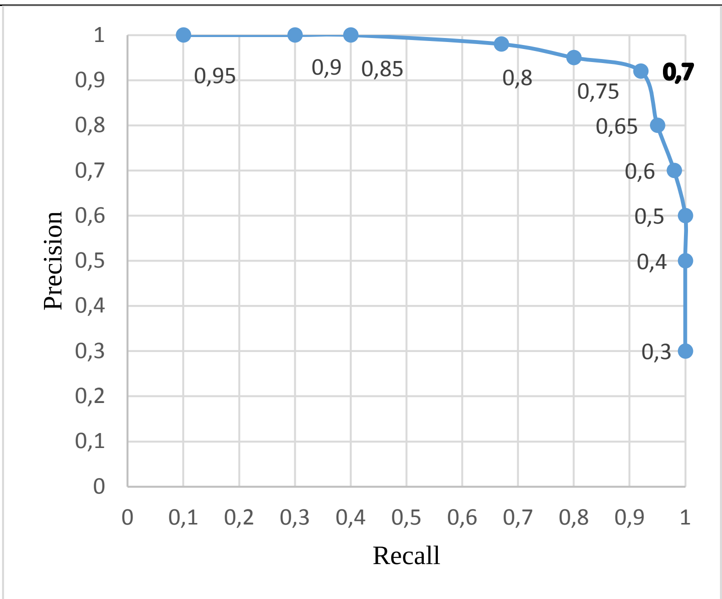

Figure 2. The curve of precision depending on the recall.

Due to the fact that video surveillance cameras are placed in various places on construction sites, and the trajectory of the workers is random, the images of workers on the frames can be of different resolution. The shooting distance is inversely proportional to the number of pixels in the image occupied by the worker's representation in the frame, as well as the number of features that can be extracted from the given image. From where we can conclude that at a long distance it is difficult to recognize the object in the image. To evaluate the method's performance with objects of different resolutions, the images of the test sample were divided into three categories depending on the distance to the target object: long, medium and short distance. 1000 images from the test dataset were selected for each category. The results are presented in table 2.

Despite the fact that as the distance between the camera and the workers increased, the classifier's performance metrics deteriorated, the overall precision of the classifier remained above 90%. Consequently, the trained model proved to be robust in detecting objects with various resolutions. Changes in weather conditions have a significant influence on the quality of classification of images obtained in open production sites. Since, in heavy rain, snow and fog work on construction sites is usually suspended, it can be excluded from consideration. Then we need to test the method on images taken under the following weather conditions: sunny, cloudy, rainy and foggy. The test results (Table 3) showed the robustness of the method to weather changes. However, the best results were obtained in sunny weather (precision — 96%), the worst in rainy (precision — 93%). To test the effect of light on the results of the method, we divided the frames from the test dataset into four categories: images taken from 8:00 to 10:00, from 10:00 to 13:00, from 13:00 to 16:00, from 16:00 to 18:00.

The test results showed that the method works robustly under different lighting conditions. From table 3 it can be seen that decrease in light, the speed, and precision of the method changed slightly.

Table 2. Performance metrics of the system depending on the distance to the object.

Category	Subcategories	Precision (%)	Recall (%)	Miss rate (%)	Time (s)
Distance	Short	98.4	95.9	4.1	0.204
	Medium	95.8	95.3	4.7	0.207
	Long	93.7	92.3	7.7	0.212

Table 3. Performance metrics of the system depending on weather conditions.

Category	Subcategories	Precision (%)	Recall (%)	Miss rate (%)	Time (s)
Weather	Sunny	96.7	95.2	4.8	0.204
	Cloudy	95.7	95.8	4.2	0.202
	Rainy	93.7	94.2	5.8	0.209
	Foggy	94.7	93.0	7.0	0.210

Table 4. Performance metrics of the system depending on the light at different times of the day.

Category	Subcategories	Precision (%)	Recall (%)	Miss rate (%)	Time (s)
Light	08-10	95.6	94.6	5.4	0.209
	10-13	96.6	95.8	4.2	0.207
	13-16	97.0	95.5	4.5	0.208
	16-18	96.9	93.7	6.3	0.210

Depending on the type, place of work and the tool used, workers have to take different poses. Thus, to test the quality of the algorithm were selected images containing workers in the positions: standing, bending, squatting and sitting. Each category was represented by 1000 images. The test result (table 5) shows high precision at various poses of workers. (more than 90%). The worst results were obtained in the squat position (93% Precision).

Table 5. Performance metrics of the system depending on poses.

Category	Subcategories	Precision (%)	Recall (%)	Miss rate (%)	Time (s)
Poses	Standing	96.8	96.9	3.1	0.209
	Bending	95.6	94.0	6.0	0.208
	Squatting	93.7	93.5	6.5	0.205
	Sitting	94.6	98.4	1.6	0.207

Construction sites are usually occupied by many workers, equipment and building materials. On the frames obtained from CCTV, the images of the workers were often overlapped by various obstacles: equipment, building materials, etc. Therefore, on many frames, workers are only partially visible. To test the effect of partial overlaps, we classify the degree of overlap into four categories: “whole body visible”, “upper body visible”, “head visible”, and “part of head visible” (Table 6). The test results showed that precision for the first three categories was more than 95%. For part of the heads, the detection precision was 90%, but the recall is only 64%.

Table 6. Performance metrics of the system depending on the presence of overlaps.

Category	Subcategories	Precision (%)	Recall (%)	Miss rate (%)	Time (s)
Overlap	Whole body visible	95.5	95.3	4.8	0.205
	Upper body visible	96.2	97.8	2.2	0.206
	Head visible	96.0	95.2	4.8	0.204
	Part of head visible	90.1	61.3	38.8	0.209

5. The discussion of the results

The article proposes a new method for detecting workers without hardhats on construction sites. The proposed method can monitor in real time with high accuracy and robustness to various scene changes. Thus, this method can provide early information about the absence of PPE on the worker while on the production site. This article discusses various methods for detecting hard hats and object detection technologies in general. Previous studies have used methods to solve this problem with limitations on robustness and practical applicability in the conditions of the open production site. We looked at the limitations of each of these methods and discussed the development of vision-based methods in the history of computer vision. Existing sensor-based detection methods, including RFID-based methods, are limited by the need for a physical tag or sensor that each builder must wear. Sensors that transmit data via Bluetooth require regular charging, which affects the practicality of its use in an open production site. In addition, these systems are more dependent on the actions of employees, and their implementation requires large investments. Given the limitations of HOG in practical use and the high efficiency of methods using deep learning, the proposed method can significantly improve the quality of automatic detection of the absence of hard hats on workers. When choosing a method for use on construction sites, we analyzed the characteristics of images, open production sites, and a number of factors that affect the detection of the absence of hard hats. The Faster R-CNN algorithm was chosen as this method. To test the robustness of the method, we tested the work of the Faster R-CNN on various images of the construction site. As a test data set, 20,000 images were collected. The test data set covers a variety of visual changes that can occur on outdoor production sites, including weather, lighting, worker poses, and overlaps. The results obtained in the course of experimental testing prove the robustly of the Faster R-CNN method for revealing the absence of hard hats in various visual conditions. Recognition precision and recall in all cases exceeded 90%, except low recall for the case of visibility of the upper part of the head — which is the expected result, because even none of the previously known algorithms does give an accurate result in this case. However, given the ability of the Faster R-CNN operate in real time and frequent changes in the poses of workers, we can assume that with a high probability the image of the worker will get completely into the frame at the next moment of time, where the head will be visible.

6. Conclusion

Construction remains one of the most dangerous employment sectors in the world. Despite the fact that hard hats provide significant protection from falling objects and from blows to the head, they do not always prevent accidents on the spot, resulting to head injury. For effective safety management at production sites, it is essential to improve the monitoring of workers without hard hats. This paper proposes a method for detecting the absence of hard hats, capable of operating in real time in changing conditions of open production sites, based on the neural network method of classification of objects Faster R-CNN. The results of testing the method showed that the proposed method was able to successfully detect workers without hard hats under different conditions with an average precision and recall of 95.7% and 94.9%, respectively. High metrics of precision and recall show that the proposed method can be effectively used in video surveillance systems to detect workers without hard hats in real time.

7. References

[1] Schneider S, Susi P 1994 Ergonomics and construction: a review of potential hazards in new construction *Am. Ind. Hyg. Assoc. J* **55(7)** 635-649

[2] Jeong B Y 1998 Occupational deaths and injuries in the construction industry *Appl. Ergon.* **29(5)** 355-360

[3] Park M-W, Elsafty N and Zhu Z 2015 Hardhat-wearing detection for enhancing on-site safety of construction workers *J. Constr. Eng. Manag.* **141(9)** 04015024

[4] Paul M, Haque S M and Chakraborty S 2013 Human detection in surveillance videos and its applications-a review *EURASIP J. Adv. Signal Process* **1** 176

[5] Tian Y, Feris R S, Liu H, Hampapur A and Sun M-T 2011 Robust detection of abandoned and removed objects in complex surveillance videos *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **41(5)** 565-576

[6] Huang X, Hinze J 2003 Analysis of construction worker fall accidents *J. Constr. Eng. Manag.* **129(3)** 262-271

[7] Kelm A, Lombardi D A, Verma S K, Brennan M J and Perry M J 2013 Factors influencing worker use of personal protective eyewear *Accid. Anal. Prev.* **41(4)** 755-762

[8] Flin R, Mearns K, O'Connor P and Bryden R 2000 Measuring safety climate: identifying the common features *Saf. Sci.* **34(1)** 177-192

[9] Barro-Torres S, Fernández-Caramés T M, Pérez-Iglesias H J and Escudero C J 2012 Real-time personal protective equipment monitoring system *Comput. Commun.* **36(1)** 42-50

[10] Dong S, He Q, Li H and Yin Q 2015 Automated PPE Misuse Identification and Assessment for Safety Performance Enhancement *ICCREM* 204-214

[11] Han S, Lee S 2013 A vision-based motion capture and recognition framework for behavior-based safety management *Autom. Constr.* **35** 131-141

[12] Han S, Lee S and Peña-Mora F 2013 Comparative study of motion features for similarity-based modeling and classification of unsafe actions in construction *J. Comput. Civ. Eng.* **28(5)** A4014005

[13] Ray S J, Teizer J 2012 Real-time construction worker posture analysis for ergonomics training *Adv. Eng. Inform.* **26(2)** 439-455

[14] Starbuck R, Seo J, Han S and Lee S 2014 A stereo vision-based approach to marker-less motion capture for on-site kinematic modeling of construction worker tasks *Comput. Civ. Build. Eng.* 1094-1101

[15] Liu M, Hong D, Han S and Lee S 2016 Silhouette-Based On-Site Human Action Recognition in Single-View Video *Construction Research Congress* 951-959

[16] Du S, Shehata M and Badawy W 2011 Hard hat detection in video sequences based on face features, motion and color information *3rd International Conference on Computer Research and Development* **4** 25-29

[17] Rubaiyat A H, Toma T T, Kalantari-Khandani M, Rahman S A, Chen L, Ye Y and Pan C S 2016 Automatic Detection of Helmet Uses for Construction Safety *WIC/ACM International Conference on Web Intelligence Workshops (WIW)* 135-142

[18] Myasnikov V V 2017 Description of images using model-oriented descriptors *Computer Optics* **41(6)** 888-896 DOI: 10.18287/2412-6179-2017-41-6-888-896

[19] Ren S, He K, Girshick R and Sun J 2015 Faster R-CNN: towards real-time object detection with region proposal networks *Adv. Neural Inf. Process. Syst.* 91-99

[20] PASCAL VOC 2017 *Detection results: VOC2012*

[21] Ott P, Everingham M 2009 Implicit Color Segmentation Features for Pedestrian and Object Detection *12th International Conference on Computer Vision* 723-730