AMERICAN UNIVERSITY OF BEIRUT

AUTOMATED VISION-BASED HARDHAT-WEARING DETECTION SYSTEM FOR CONSTRUCTION SAFETY APPLICATIONS

by
BAHAA EDDINE MNEYMNEH

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Engineering
to the Department of Civil and Environmental Engineering
of the Faculty of Engineering and Architecture
at the American University of Beirut

Beirut, Lebanon
May 2017

AMERICAN UNIVERSITY OF BEIRUT

AUTOMATED VISION-BASED HARDHAT-WEARING DETECTION SYSTEM FOR CONSTRUCTION SAFETY APPLICATIONS

by

BAHAA EDDINE MNEYMNEH

Approved by:

Dr. Hiam Khoury, Associate Professor Civil and Environmental Engineering	Advisor
Dr. Farook Hamzeh, Assistant Professor Civil and Environmental Engineering	Member of Committee
Dr. Ibrahim Alameddine, Assistant Professor Civil and Environmental Engineering	Member of Committee

Date of thesis defense: April 21, 2017

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT RELEASE FORM

Student Name:

____

Last First Middle

○ Master's Thesis ○ Master's Project ○ Doctoral Dissertation

☐ I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

☐ I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes after:

One ---- year from the date of submission of my thesis, dissertation, or project.
Two ---- years from the date of submission of my thesis, dissertation, or project.
Three ---- years from the date of submission of my thesis, dissertation, or project.

____

Signature Date

ACKNOWLEDGEMENTS

I am thankful to ALLAH for all his blessings, for the good health and for the well-being that were necessary to complete this thesis.

I am using this opportunity to express my gratitude to everyone who supported me throughout the course of my study at the American University of Beirut.

First and foremost, special thanks to my parents for their unconditional love and support throughout my academic career. Thank you both for giving me strength to chase and reach my dreams.

I wish to express my sincere thanks to my advisor, Dr. Hiam Khoury for her inspiring guidance and constructive criticism during the thesis work. I am sincerely grateful to her for sharing her truthful and illuminating views and experiences on several issues related to this thesis.

I also take this opportunity to express gratitude to all of the members of the Civil and Environmental Engineering department for their help and support.

AN ABSTRACT OF THE THESIS OF

Bahaa Eddine Mneymneh for Master of Engineering
Major: Construction Engineering and Management

Title: Automated vision-based hardhat-wearing detection system for construction safety applications

The construction industry is still considered among the most dangerous industries in the world as workers are exposed to a constant risk of getting injured from falls, slips, trips, or getting struck by moving or falling objects. In an attempt to provide a safe working environment, safety programs have been trying to impose an array of provisions and regulations, of which enforcing the use of personal protective equipment (PPE), in particular hardhats, proved to be of paramount importance. As the awareness and perception of construction workers on safety and hardhat use cannot be fully trusted, the responsibility has been traditionally put into the hands of safety officers to ensure compliance with these safety regulations. However, the task of actively supervising a large construction site with a sizeable number of workers is considered tedious, inefficient, and time-consuming. Hence, this research aims at creating a vision-based system that can automatically detect a failure to wear the hardhat in videos captured from construction sites. The objective of this study is thereby two-fold: (1) evaluating existing computer vision techniques in efficiently detecting hardhats on jobsites, and (2) developing an integrated vision-based framework that can actively identify mobile construction workers then search for the presence of a hardhat in the upper region of the detected personnel. Components of the complete framework were implemented and results highlighted the potential of the proposed automated hardhat detection system in enhancing construction safety inspections.

CONTENTS

ILLUSTRATIONS

TABLES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Construction is still considered one of the most dangerous industries despite numerous measures and precautions aiming at improving the overall safety on construction sites. 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 while the mining sector caused 183 (United States Department of Labor 2014). Causes of construction worker deaths include falls, being struck by objects and caught-in/between, that contribute to over 50% of the total fatalities in the industry (United States Department of Labor 2014). Wearing Personal protective equipment (PPE) and more specifically hardhats and industrial safety helmets reduces the risk of injury by impact from falling or flying objects (HSE 2016). Hence, safety inspections are regularly performed to maintain a safe working environment. These commonly consist of safety officers moving around the construction site to uncover unsafe working practices, and establish conformity with health and safety requirements. However, the duty of actively supervising a sizable number of workers and constantly identifying all potential violations is still rated as manual and laborious (Ham et al. 2016). Moreover, especially in developing countries, the construction labor safety laws are not properly enforced (Teo et al. 2008, and Chiocha et al. 2011). Even on projects where safety programs are initiated, proper observations and follow up are rare (Recate Suazo and Jaselskis 1993). In a recent study, contractors stated that not having a safety and health management system (SHMS) on their construction sites was caused by the absence of law requirements, the high cost of safety implementation and its time consumption (Awwad et al. 2016). Hence, automating and optimizing the current safety procedures would offer contractors more incentives to implement it, and will ultimately provide more protection to construction workers.

1.2. Worker perception towards PPE

To further assess the importance of properly enforcing the use of personal protective equipment on construction sites, a survey involving 285 construction related personnel of various positions and years of experience, from 7 different construction sites across Lebanon, was conducted. As part of the evaluation of their perception on hardhats use, interviewees were asked to rate the importance of hardhats on a scale from 1 to 10 and whether the reason behind wearing it was out of personal incentive or it was imposed by safety officers. The survey includes as well a question on past incidents, whereby the answer “Yes” indicates that the construction personnel or any of the co-workers, have previously been protected from any sort of minor or major head injury by wearing a hardhat. The summary of results is presented in Table 1.1.

Table 1.1: Relationship between hardhat use and past incidents

		Reason to wear a hardhat		
		Personal incentive	Imposed by site regulators	**Total**
Past incident	Yes	62	16	**78**
	No	71	136	**207**
	Total	**133**	**152**	**285**

According to table 1, the following findings can be formulated:

- 27% of all interviewees had previous incidents involving hardhats and 73% did not.
- 47% of the total interviewees would wear a hardhat out of personal incentive while 53% would not.
- 79% of interviewees who had a past incident would wear a hardhat out of personal incentive and 65% who have not experienced a past incident would only wear a hardhat whenever it was imposed by safety officers.

The last finding clearly implies that the use of hardhat is significantly higher for workers that have survived an incident despite it being a common safety practice. This is highlighted through the average hardhat score that was computed as 4.88 in the case of individuals who have not previously experienced any incident compared to 7.55 for individuals who have experienced a past incident (Figure 1.1).

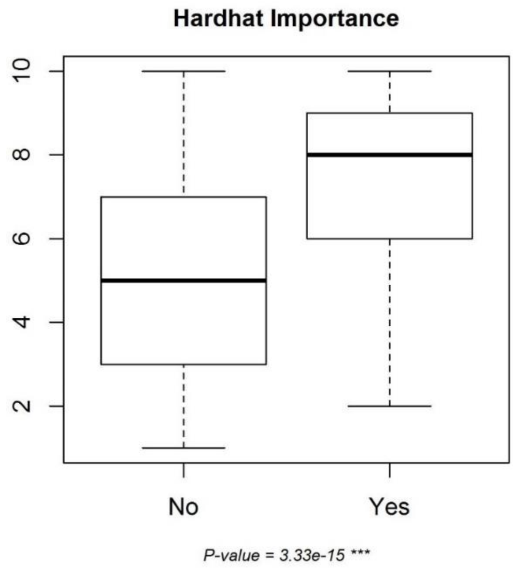

Figure 1.1: Impact of past incidents on hardhat evaluation

1.3. Objective and thesis structure

Because the workers' awareness could not be trusted to fully comply with safety requirements and regulations, and since safety officers are often incapable of actively monitoring large construction areas, this research aims to create an automated tool that can accurately detect any failure in wearing personal protective equipment (PPE), and more specifically hardhats. The objective of this study is thereby two-fold: (1) evaluating existing computer vision techniques in efficiently detecting hardhats on jobsites, and (2) developing an integrated vision-based framework that can actively identify mobile construction workers then search for the presence of a hardhat in the upper region of the detected personnel. Separate components were created and the complete framework was later implemented.

This thesis presents the work achieved towards the realization of the aforementioned objectives, and consists, besides this introductory chapter, of two stand-alone research papers:

Paper 1 evaluates existing computer vision techniques, in particular object detection methods, with the aim of identifying the most suitable approach for fast and accurate hardhat detection in construction environments.

Paper 2 presents the design, implementation, and assessment of a new integrated and complete framework that: (1) presents a novel vision-based motion detection method based on standard deviation differences between consecutive images to detect mobile construction workers, and (2) relies on a cascade object detector coupled with color analysis and segmentation to detect hardhats in the upper region of the detected personnel with high values of precision and recall.

Relevant background and literature are presented withing the scope of each paper.

References

Awaad, R., El Souki, O., and Jabbour, M. (2016). Construction safety practices and challenges in a Middle Eastern developing country. *Safety Science,* vol. 83, 1-11.

Ham, Y., Han, K. K., Lin, J. J. and 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*, vol. 4.

Health and Safety Executive (2016). Personal protective equipment (PPE). [Online]. Available: http://www.hse.gov.uk/toolbox/ppe.htm. [Accessed 2016 October 2016].

Recate Suazo G. A., and Jaselskis, E. J. (1993). Comparison of Construction Safety Codes in United States and Honduras. *Journal of Construction Engineering and Management,* 119 (3), 560-572.

United States Department of Labor, (2014). Commonly Used Statistics. [Online]. Available: https://www.osha.gov/oshstats/commonstats.html. [Accessed 8 October 2016].

United States Department of Labor, (2014). Fatal occupational injuries by industry and event or exposure, all United States, 2014. [Online]. Available: http://www.bls.gov/iif/oshwc/cfoi/cftb0286.pdf. [Accessed 8 October 2016]

CHAPTER 2

EVALUATION OF COMPUTER VISION TECHNIQUES FOR AUTOMATED HARDHAT DETECTION IN CONSTRUCTION SAFETY APPLICATIONS

Abstract

Despite various safety inspections carried out over the years to ensure compliance with regulations and maintain acceptable and safe working conditions, construction is still among the most dangerous industries responsible for a large portion of the total worker fatalities. A construction worker has a chance of 1-in-200 of dying on the job during a 45-year career, mainly due to fires, falls, and being struck or caught-in/between objects. This in part can be attributed to how monitoring the presence and proper use of personal protective equipment (PPE) on jobsites by safety officers becomes inefficient when surveying large areas and a considerable number of workers. Therefore, this paper takes the initial steps and aims at evaluating existing computer vision techniques, namely object detection methods, in rapidly detecting, from videos captured on many indoor jobsites, whether workers are wearing hardhats. Experiments have been conducted and results highlighted the potential of cascade classifiers, in particular, in accurately and precisely detecting hardhats under various scenarios and for repetitive runs.

Keywords

Computer vision, Hardhat, Construction safety, feature detection, template matching, cascade classifier

2.1. Introduction

A construction worker has a chance of 1-in-200 of dying on the job during a 45-year career. This statistic reveals the extent of hazard a construction laborer is exposed to on a daily basis. According to the Unites States Department of Labor, 874 construction workers were killed while performing job related activities in 2014. The health and safety executive states that a total of 1.7 million working days were lost in the United Kingdom in 2014, as 6% of workers sustained or suffered from a work-related injury or illness. A worker's death not only constitutes a tragic event to his family and surrounding, but may also provoke major legal consequences and project delays. Non-fatal injuries and illnesses are also a concern to construction parties as they incur significant waste of time and money. As such, proper use of personal protective equipment (PPE) greatly reduces the risk of injury. The hardhat is considered as one of the most fundamental means of protection against head injuries from impact with flying or falling objects. The importance of safety helmets is highlighted in the widely-adopted safety codes and regulations, which impose wearing a protective hat on any person entering a construction environment. However, the behavior and awareness of construction workers towards the importance of safety equipment cannot be fully trusted. A large portion of work related head injuries is sustained by workers not wearing hardhats. In the United States, according to the bureau of labor statistics, 84% of workers injured in the head region were not utilizing any form of acceptable protection (OSHA 2006). For this reason, safety personnel are typically deployed on construction sites to ensure compliance with safety regulations and maintain acceptable working conditions. Automating the current time-consuming safety procedures would create a safer environment where no preventable life loss occurs. Therefore, in this paper, existing computer vision techniques are evaluated in rapidly and automatically detecting hardhat use. In order to test the system's ability to detect hardhats in dynamic indoor construction environments, the methods are tested against numerous situations, including variations in orientations, colors, background contrast, image resolution, and lighting conditions.

2.2. Background

2.2.1. Use of information technology in construction safety

The construction field is starting to catch up with other fields regarding the use of information technology tools. In recent years, research has been carried out in an attempt to improve and automate the construction process in many different areas, including on-site safety whereby safety programs implemented proved to be manual, tedious and time-consuming, Skibniewski (2014) reported that 136 articles focusing on IT applications in this field were published between 2006 and 2014. Research topics include sensor-based systems, robotics and manipulators, and information analysis, management and reporting. Sensor based systems in the field of construction safety rely on a wide range of “Real Time Locating Systems” (RTLS). Commonly used sensors include “Radio Frequency Identification” (RFID) sensors, “Ultra-wide Band” (UWB) sensors, and laser sensors. RFID sensors are mainly employed to develop early warning systems for workers. Yoshida and Chae (2010) utilized RFID sensors to develop a system that informs workers about potential safety hazards on site. Their system prevents collision accidents by warning workers and equipment operators entering a possible conflict zone, using real-time localization. Marks and Teizer (2013) evaluated the potential use of semi-passive RFID in a pro-active real time personal protection unit (PPU), that protects workers against being struck by a machine or equipment. Their tests revealed that a PPU system based on semi passive two-dimensional RFID requires covering all possible worker-to-machine positions and orientations to be effective. They concluded that further research is required for the development of such PPU systems, particularly regarding the use of three-dimensional RFID. Wireless sensors provide real-time localization of workers, material and equipment. Several studies about the accuracy and reliability of UWB sensors for safety applications have been performed

(Teizer and Castro-Lacouture 2007, Hwang 2012). Acceptable accuracy levels could be obtained in sizeable open space construction sites (Cheng et al. 2011). Laser sensors and scanners are used to examine the visibility of a tower crane or equipment operator, detect his blind spots, and provide information about potential hazards to his surrounding workers (Lee et al. 2012, and Cheng and Teizer 2014). On the other hand, robotics and manipulators are replacing human workforce with robots for tasks performed under dangerous conditions. Kim et al. (2009) evaluated the performance of the Hume Concrete Pipe Manipulator (HCPM), a system commissioned by the Korea Ministry of Construction and Transport to perform pipe laying work. This task previously required laborers to work in trenches and was considered one of the most dangerous construction operations. In addition to eliminating potential safety hazards, analysis of the HCPM's performance indicated a 65% improvement in productivity and a 33% reduction is costs. Other research efforts resorted to information analysis, management and reporting by adopting BIM for safety planning and schedule-workspace interference visualization, to secure work performance safety and decrease collision risk between resources. Ding et al. (2012) created a safety risk identification system (SRIS) that automatically identifies hazards related to metro and underground construction using construction drawings. However, none of the aforementioned studies used computer vision in automating safety.applications.

2.2.2. Use of computer vision in construction

Computer vision techniques have been applied to extract relevant information from construction sites. It is used to automate field observations and replace present time-consuming data collection methods. Applications include progress monitoring, productivity analysis, object detection, tracking and motion detection, and action recognition (Yang et al. 2015). Computer vision-based systems have also found a direct application in construction safety, in identifying unsafe acts, hazardous situations and awkward postures (Seo et al. 2015). For instance, texture recognition systems have been used to identify on-site material for construction progress monitoring (Liu et al. 2010). Those systems, however, have only been tested under controlled conditions and their accuracy drops significantly when used with images from construction sites. To address this issue, Dimitrov and Golparvar-Fard (2014) proposed a new vision-based algorithm that classifies material using joint probability distribution of responses. The system was capable of accurately classifying material from images taken from various viewpoints and under different conditions of illumination. Hamledari et al. (2016) presented a computer vision-based system that relies on shape and color analysis, to detect indoor partition components such as studs, insulation, electrical outlets and drywall sheets. Based on detection results from images and videos from construction sites, the system classifies the current stage of work into one of five possible stages: (1) framing, (2) insulation, (3) installed drywall, (4) plastered drywall, and (5) painted partition. The obtained information may then be utilized to update corresponding BIM models and schedules. Additionally, vision-based tracking frameworks are used to address the limitations of using sensor-based systems, such as high costs, time consumption and privacy issues. Those systems use data from multiple static cameras to identify, track and calculate 3D coordinates of construction entities from each frame (Brilakis et al. 2011, and Park et al. 2012a). Chi et al. (2009) proposed an algorithm that identifies and tracks objects on construction sites using a high-frame-rate range sensor. Image matching algorithms were used to classify objects using a model database and through comparison with previous scans. Memarzadeh et al. (2013) presented a computer vision-based framework to automatically detect construction workers and equipment using videos streamed from construction sites. They utilized an algorithm based on Histogram of Oriented Gradients and Colors (HOG + C) to make the semi-automated detection less time-consuming. Motion detection systems generally rely on background subtraction algorithms to detect foreground and moving objects (Park et al. (2015). Park et al. (2012b) utilized standard background subtraction to detect moving blobs in an image. Within the moving foreground, HOG features are used to identify peoplethen color gradients distinguish between pedestrians and construction workers, based on the color of the safety vest.

As computer vision proved promising in the field of construction, and safety in particular, recent studies have adopted it for automating hardhat detection in an attempt to efficiently and rapidly protect construction personnel against head injuries and collisions.

2.2.3. Use of computer vision for hardhat detection

Proper use of personal protective equipment is essential to provide construction workers with adequate levels of safety. Research has been recently carried out to facilitate, improve, and automate appropriate PPE usage enforcement procedures (Bajracharya 2013). 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 was detected using existing face detection algorithms based on Haar-Like features, introduced by Lienhart and Maydt (2002) and Viola and Jones (2004). The system than 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 not real-case scenarios from construction sites. On the other hand, Gheisari et al. (2014) evaluated the potential applications of unmanned aerial systems (UAS) in construction safety by providing safety personnel with real-time visual access to jobsites. Their system was evaluated based on the ability of the test subjects to detect workers not wearing hard-hats using images and videos captured by a quadcopter from a construction site. Results revealed that using unmanned aerial systems would be useful for safety related tasks. However, a large-size interface on a tablet device is required to provide a precise view of the jobsite. The detection process was still manual and required the safety officer's presence and perception. Shrestha et al. (2015) proposed an algorithm that detects workers using standard face detection then applied the edge detection technique on the region directly above the worker's head. The system detects a hard-hat if its outline is a semicircle and its color is red. However, their program 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. No validation of the efficiency of their system in an actual construction environment was realized. A recent study by Park et al. (2015) detected hard-hats using a support vector machine (SVM) classifier, as part of a complete framework that aims at enhancing on-site safety conditions. The algorithm is based on shape recognition and utilizes histogram of oriented gradient (HOG) features to describe the cap style shape of the hard-hat. While the proposed framework was capable of detecting a hardhat under various conditions and independently from its color, it was also susceptible to false detection because the semi-circular shape of the hardhat could easily be extracted from other irrelevant objects. As such, Rubaiyat et al. (2016) combined the use of Circle Hough Transform (CHT) feature extraction that can identify the circular shape of the hardhat, with color analysis, to create a more efficient helmet detection system. Their method was capable of detecting hardhats at a rate of 79.1%.

2.3. Limitations of existing studies and contributions

However, the work of the aforementioned research efforts is still in its infancy. The exisiting systems were either only able to detect frontal views of hard-hats under laboratory conditions and never tested under proper site conditions, or were victims of overprediction and identifying unwanted objects as hardhats. Moreover, almost all studies in this domain put too much emphasis on the values of precision (i.e. percentage of detections that are true positives) and recall (percentage of true positives detected), calculated based on test results from a large sample of random images in a database. The two parameters, often considered in describing the precision of a classification model, may not provide an accurate evaluation of the effectiveness of the proposed algorithms. This is mainly due to the fact that the performance of the detection model is not merely random and cannot be evaluated using random samples. For instance, a classifier capable of detecting hardhats based on shape recognition and strong contrast with the background would correctly identify all images when these conditions are satisfied, but would also fail to identify all images with insufficient background contrast. The calculated values of precision and recall are directly related to the nature and the level of challenge of the testing images. A more convenient approach is to test the detector against selected scenarios of various orientations, colors, background contrast, image resolution, and lighting conditions in order to qualitatively assess its effectiveness. In addition, none of the previous studies were concerned with the time efficiency of the detection method. Dealing with the safety of construction workers and personnel has to happen relatively quickly and in real-time. After all, the detection speed of the algorithm is as important as its accuracy. Therefore, the aim of this paper is to evaluate existing computer vision techniques, in particular object detection and recognition methods, in order to identify the most suitable algorithm for efficiently and rapidly detecting hardhat wearing under various indoor site conditions. The accuracy as well as time efficiency of the different vision techniques aretested in numerous situations, under various conditions and for several runs.

2.4. Methodology

This section describes and evaluates various computer vision algorithms deemed useful for detecting hardhats. Among several existing computer vision techniques, object detection/recognition methodsproved 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 the form, color, repeatability, shape, and scale variance of the target object. The components of the evaluated algorithms were implemented using Matlab 2016a.

2.4.1. Feature detection, extraction, and matching

In image recognition, a local feature is defined as a “pattern which differs from its immediate neighborhood” (Mikolajczyk and Tuytelaars 2008). Local features and descriptors are the cornerstone of a large array of computer vision techniques and applications, including object detection, tracking and motion estimation (Dickscheid et al. 2011). In the feature detection stage, feature detectors or descriptors are used and more specifically gradient-based features such as the Speeded-Up Robust Features (SURF), or binary features including Binary Robust Invariant Scalable Keypoints (BRISK) and Features from Accelerated Segment Test (FAST) are commonly used to find point correspondences between the input image, and a reference image containing the object, or objects, of interest (Ahonen et al. 2006, and Dalal and Triggs 2005, Rosten and Drummond 2006, Bay et al. 2008, and Leutenegger et al. 2011). As such, SURF, BRISK, and FAST features were considered in this study as depicted in the code snippet below:

```
Surf_Features = detectSURFFeatures(rgb2gray(Image));
Brisk_Features = detectBRISKFeatures(rgb2gray(Image));
Fast_Features = detectFASTFeatures(rgb2gray(Image));
```

On the other hand, the feature extraction locates the detected features within each image, while feature matching identifies similarities between the sample and test images, using the following code snippet based on SURF features:

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

Outliers were then removed and the transformation matrix was calculated, using Random Sample Consensus (RANSAC) algorithm (Khoury et al. 2015). A hardhat is detected when the number of matched features between the input and reference images is sufficient. The number of hardhats is then computed by hiding identified ones from the target image so that the next best match 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 works best for objects displaying non-repeating texture patterns to allow unique and numerous feature matches.

2.4.2. Template matching

Template matching often refers to a series of operations aiming at detecting and identifying a certain form or pattern in an input image, by comparison with a template image (Brunelli 2009). The concept of template matching is fairly straightforward: The template is positioned over the input image at every possible position and a similarity coefficient is calculated based on pixel values. Possible metrics to determine the similarity include the sum of absolute differences (SAD), the sum of squared differences (SSD), and the maximum absolute difference (MaxAD) (Yu et al. 2006). Template matching is best used in localizing an existing part of an overall image, and was found useful in applications such as quality control and mobile robot navigation (Aksoy et al. 2004, and Kyriacou eh al. 2005). In Matlab, 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 but more computationally expensive, while the latter is quicker, but may not always find the optimal solution. In Matlab, a template matcher is typically based on SAD unless otherwise stated (e.g. Three-Step) as shown in the code snippet below:

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

A hardhat is detected when the calculated minimum difference between a reference image containing a hardhat and the input image is less than a required threshold. Generally, template matching algorithms are limited by the available computation power due to a required high detection accuracy that necessitates lengthy iteration processes.

2.4.3. Cascade classifier

In this study, cascade classifiers 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, while negative instancesare images not containing the relevant object. Positive images are utilized by the detector to describe the shape and features of the relevant object. A varied sample of 75 positive and 164 negative instances was collected from construction and was used to train the three cascade object detectors. The region of interest (ROI) containing the hardhat needs to be selected for every image in the database, using the 'Training Image Labeler' in Matlab (Figure 2.1).

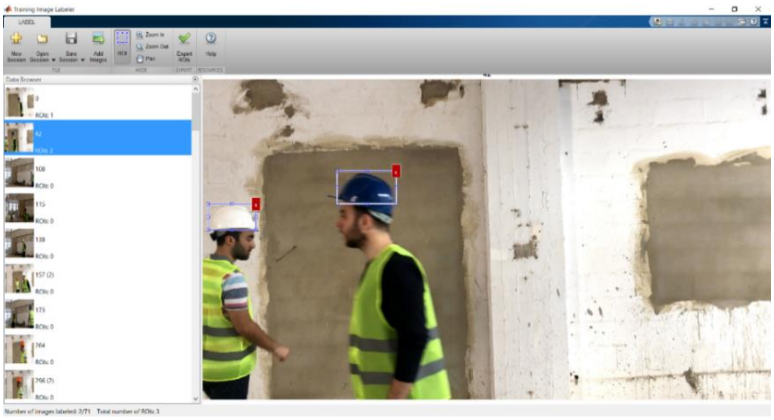

Figure 2.1: Example of ROI selection in Matlab

Once the sets of positive and negative instances are obtained, training a cascade object detector can be performed. The training process requires as well a set of input parameters, including the number of cascade stages, the true positive rate and the false alarm rate. Experimenting with those 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')
```

Relevant information about the object to be classified are stored in the created XML file.

2.5. Preliminary results and analysis

The performance of each method is assessed against variations of orientation, color, background contrast, image resolution, and lighting conditions in addition to assessing the time efficiency of each. All the required codes and algorithms were implemented in Matlab 2016a, using built-in toolboxes. Since time efficiency is a main component in our study, all experiments were carried out on the same laptop equipped with an intel i7 6700HQ Skylake processor and 16 GB of RAM.

2.5.1. Performance of the feature detection, extraction, and matching algorithm

Due to the uniform shape and color of a hardhat, the number of detected features was found to be low (Figure 2.2 and 2.3). One suggested solution to this problem was to add a customized sticker to the hardhat (Figure 2.4 a). This then greatly increased the number of extracted features (Figure 2.4 b, 2.4 c, 2.4 d).

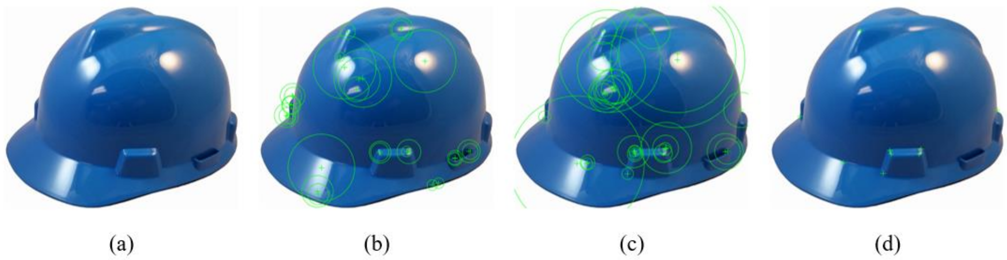

Figure 2.2: Blue Hardhat: (a) Original Image, (b) Detected SURF features (29), (c) Detected BRISK features (29), and (d) Detected FAST features (9)

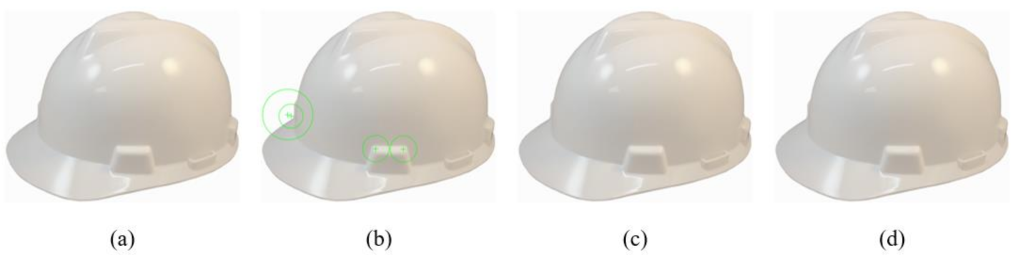

Figure 2.3: White Hardhat: (a) Original Image, (b) Detected SURF features (4), (c) Detected BRISK features (0), and (d) Detected FAST features (0)


(a) (b) (c) (d)

Figure 2.4: Hardhat with sticker: (a) Original Image, (b) Detected SURF features (285), (c) Detected BRISK features (172), and (d) Detected FAST features (424)

In order to demonstrate the ability of the local features based algorithm and evaluate its performance, the first experiment was conducted using a close-up top view image captured on a construction site, clearly showing the customized stickers on both hardhats. The algorithm is independent from any type of feature used, but given that a minimum number of features needs to be extracted with the least computational power, the choice landed on SURF features. In the first iteration, 63 matching features were found between the reference and the target image (Figure 2.5). The first detected hardhat was then hidden from the target image and in the second iteration, 44 matching features were found between the reference and the new target image (Figure 2.6).

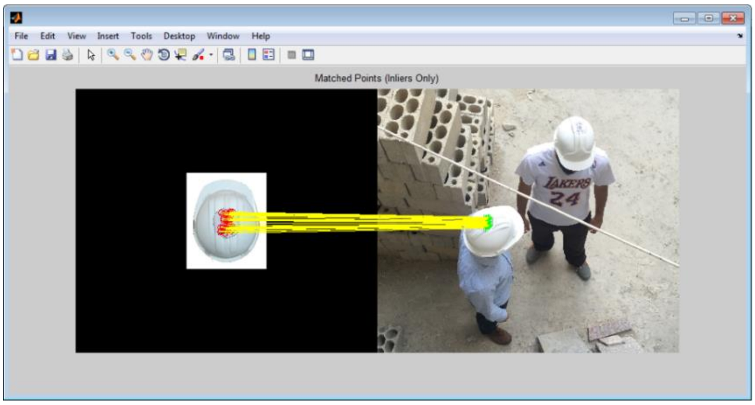

Figure 2.5: Matching features for the first hardhat

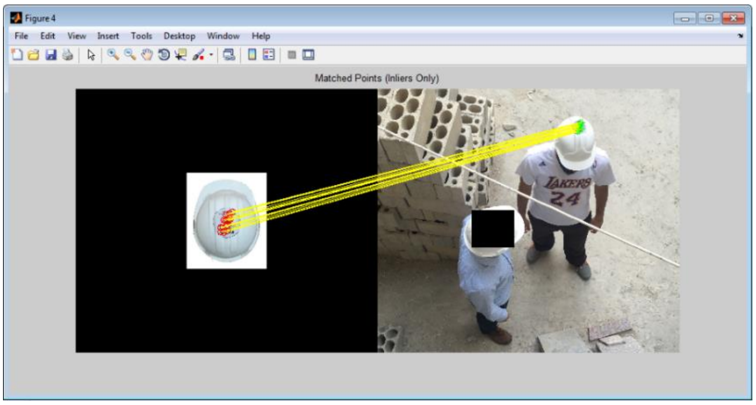

Figure 2.6: Matching features for the second hardhat

The iteration process then halts once the second hardhat is hidden and the program returns the number of detected hardhats.

Because of the lack of pertinent features on the hardhat, the algorithm searches for the customized sticker and identifies its target irrespective of the color or shape of the hardhat. However, further testing revealed some deficiencies in the system. The method is actually susceptible to misclassifying any object carrying the sticker. Moreover, in a three-dimensional dynamic construction environment, a clear view of the sticker can not be always guaranteed. As a matter of fact, in another sets of experiments, a hardhat could not be detected because either the size or resolution of the sticker was low, or the sticker was not visible due to the orientation of the hardhat (Figure 2.7). Additionally, the feature extraction and filtering together with the iteration processes required a relatively high calculation cost.

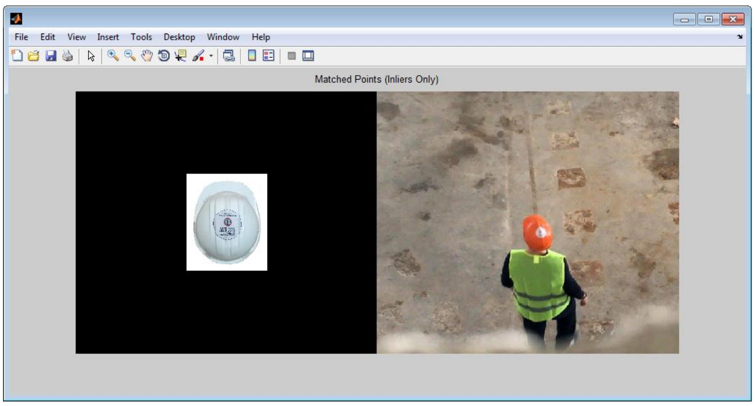

Figure 2.7: Example of no detection – Sticker size too low

2.5.2. Performance of the template matching algorithm

As shown in Figure 2.8, the algorithm wrongly predicted the location of a blue hardhat when using a template with a slightly different rotation.

Figure 2.8: Wrong detection using template matching

As such, a unified template is not sufficient to detect all instances and a classic template matching is relatively inaccurate when dealing with any form of difference in scale and rotation. Futhermore, 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 a lot of computational power

2.5.3. Performance of the cascade classifier

Cascade object detectors have proven capabilities in identifying objects of similar shapes and aspect ratios, such as face and pedestrian recognition (Lienhart and Maydt 2002, and Viola and jones 2004). Objects detectors are often sensitive to out of plane transformation. However, this should not be a problem in the case of a hardhat detection since its semi-circular shape remains unchanged regardless of the viewing angle.

Cascade classifiers based on Haar and LBP features yielded high rates of wrong detection in all testing images (Figure 2.9). However, a classifier based on HOG features provided an acceptable accuracy in this particular application. HOG features could accurately describe the circular shape of the hardhat independently from its color (Figure 2.10).

Figure 2.9: High rate of incorrect detections: (a) Using Haar Features, and (b) Using LBP Features

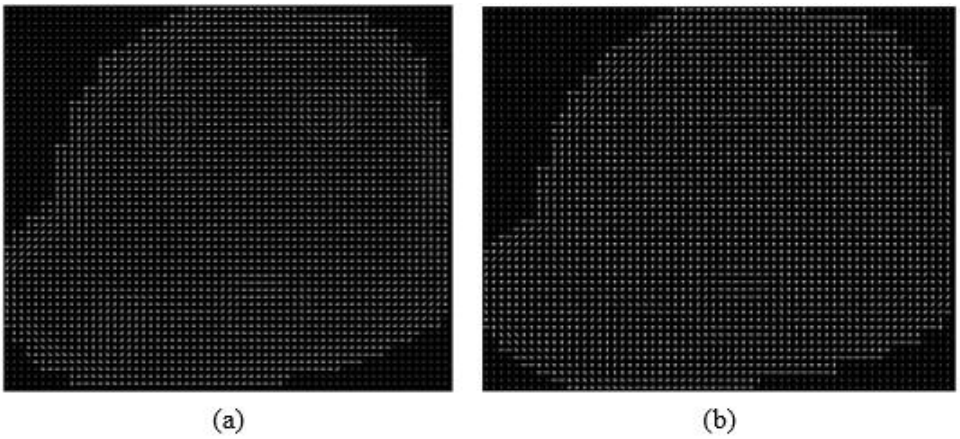

Figure 2.10: Extracted HOG features: (a) blue hardhat, and (b) white hardhat

The ability of the detector to correctly identify hardhats from different perspectives is verified as well using a large set of testing images containing front, side and back views of a blue hardhat (Figures 2.11, 2.12 and 2.13).

Figure 2.11: Detected Hardhat – Front View

Figure 2.12: Detected Hardhat – Side View

Figure 2.13: Detected Hardhat – Back View

The classifier was also capable of recognizing two objects simultaneously. Color variations also had no effect and the computational speed was acceptable. The capacity to experiment with training parameters to obtain different results is also another main advantage of such cascade classifiers.

2.5.4. Comparison and selection

Based on preliminary experiments and results featured above, each technique is assessed against the following criteria summarized in Table 2.1.

Table 2.1: Comparative summary of hardhat detection techniques

	Feature Detection, Extraction and Matching	**Template matching**	**Cascade classifier**
Computational Duration	Medium	Very High	Low
Color invariance	Yes	No	Yes
Orientation Invariance	No	No	Yes
Practicality	No	No	Yes
Customizability	Yes	No	Yes
Requires Training Database	No	No	Yes

The above comparative summary clearly reflects that the cascade object detector outperforms the other computer vision techniques and can be potentially adopted in real-time safety applications, in particular hardhat detection.

2.6. Experimental analysis of the cascade object detector

In this section, further assessment of the cascade object detector is carried out. When training a detector, a common misconception is to put a large emphasis on maximizing the size of training images. In reality, and based on our preliminary tests, better results were obtained using a compact training set of similar positive instances, and a larger generalized training set yielded unwanted results. Additionally, consistency within the images is more important than the total sample size. For this reason, all images in our obtained database have the same resolution (3840 × 2160 pixels), and the obtained detector produces optimal results when tested on images having the same resolution as database images.

Two seven-stage cascade object detectors were trained using the same image datasets of 75 positive and 164 negative images, and given two different values of false alarm rates set to 0.05 and 0.1. In theory, a larger false alarm rate should yield more false positive results, and the detector should be less likely to miss a desired object. In the following subsections, the effectiveness of the classifier is tested against variations in orientations and colors, background contrast, image resolution, and lighting conditions. The execution of the algorithm for each test scenario is timed using the "Run and Time" tool in Matlab.

2.6.1. Scenario 1: High contrast against background, variable colors and orientations:

In this scenario, the level of challenge was relatively low. All hardhats could be easily discerned from their respective backgrounds. The two detectors were tested on 10 images with 13 hardhats in total and the results are summarized in table 2.2:

Table 2.2: Performance of cascade object detectors for scenario 1

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- False alarm rate = 0.05	1	1	0	0	1	2	1	2	1	3
Detected- False alarm rate = 0.1	1	1	1	1	1	2	1	2	2	3

As expected, the cascade object detector with the lower false alarm rate missed three, while the detector with the false alarm rate of 0.1 did not miss any hardhat. Nevertheless, both classifiers were subject to wrong identification, and for instance, in image 6, an object with a similar shape as the hardhat was mistakenly classified (Figure 2.14). The time statistics of both detectors were similar with an average processing time per image of 2 seconds (tables 2.3 and 2.4).

Figure 2.14: Wrong classification of head region in image number 6 – Scenario 1

Table 2.3: Time statistics for scenario 1 – False Alarm Rate = 0.05

Action	Number of Calls	Total Time	Percentage of Total Time	Average Time
Execution of detector	10	19.949 s	91.9 %	1.994 s
Reading image file	10	1.734 s	8.0 %	0.173 s
All other actions	-	0.031 s	0.1 %	0.003 s
Total	-	21.714 s	100 %	2.171 s

Table 2.4: Time statistics for scenario 1 – False Alarm Rate = 0.1

Action	Number of Calls	Total Time	Percentage of Total Time	Average Time
Execution of detector	10	20.517 s	91.9 %	2.051 s
Reading image file	10	1.767 s	7.9 %	0.176 s
All other actions	-	0.034 s	0.2 %	0.003 s
Total	-	22.318 s	100 %	2.231 s

2.6.2. Scenario 2: Low contrast against background, variable colors and orientations

In this scenario, the level of challenge was greatly increased. The low contrast between the hardhat and its background may reduce the significance of the detected HOG features which, in turn, canreduce the efficiency of the detector. Similar to the first scenario, 10 images containing one or more hardhats were selected and results for both detectors are summarized in table 5.

Table 2.5: Performance of cascade object detectors for scenario 2

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- False alarm rate = 0.05	0	1	0	1	0	0	1	1	1	1
Detected- False alarm rate = 0.1	0	1	1	1	1	0	1	2	2	1

In this case, the performance of the cascade object detector dropped. For a false alarm rate equal to 0.1, the detector was able to identify 10 out of 13 hardhats. Nevertheless, the detector is still considered efficient even when the contrast with the background is very minimal (Figure 2.15). The time statistics of both detectors did not significantly vary and the average processing time per image was around 2 seconds (tables 6 and 7).

Figure 2.15: Correct identification of both hardhats in image 8 – Scenario 2

Table 2.6: Time statistics for scenario 2 – False Alarm Rate = 0.05

Action	Number of Calls	Total Time	Percentage of Total Time	Average Time
Execution of detector	10	20.120 s	91.8 %	2.012 s
Reading image file	10	1.765 s	8.1 %	0.176 s
All other actions	-	0.030 s	0.1 %	0.003 s
Total	-	21.915 s	100 %	2.191 s

Table 2.7: Time statistics for scenario 2 – False Alarm Rate = 0.1

Action	Number of Calls	Total Time	Percentage of Total Time	Average Time
Execution of detector	10	20.148 s	91.7 %	2.014 s
Reading image file	10	1.778 s	8.1 %	0.174 s
All other actions	-	0.036 s	0.2 %	0.003 s
Total	-	21.962 s	100 %	2.196 s

2.6.3. Scenario 3: Low luminosity

Unlike scenario 2, the variation of image luminosity did not considerably affect the performance of the cascade object detector. This is due to the fact that the HOG features are capable of describing the shape of the object irrespectiveof its color.

2.6.4. Scenario 4: Different image resolutions

To test the effect of changing the resolution of test images independently from other factors, this experiment was carried out using the same images of scenario 1, cropped or resized to obtain images having a lower resolution resolution (1920 × 1080 pixels). Images 1 to 5 were cropped, while images 6 to 10 were resized. Cropped images yielded results identical to scenario 1. On the other hand, resizing the image can possibly decrease the size of hardhats below the trained size and accordingly does not allow hardhat detection (table 8). Therefore, training the detector using images of the same resolution and taken from a similar distance as expected subject images is of paramount importance. Accordingly, the time statistics significantly improved (tables 9 and 10).

Table 2.8: Performance of cascade object detectors for scenario 4

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-False alarm rate = 0.05	1	1	0	0	1	2	0	1	1	3
Detected-False alarm rate = 0.1	1	1	1	1	1	2	0	1	2	3

Table 2.9: Time statistics for scenario 4 – False Alarm Rate = 0.05

Action	Number of Calls	Total Time	Percentage of Total Time	Average Time
Execution of detector	10	4.887 s	90.3 %	0.488 s
Reading image file	10	0.486 s	9.0 %	0.048 s
All other actions	-	0.040 s	0.7 %	0.004 s
Total	-	5.413 s	100 %	0.541 s

Table 2.10: Time statistics for scenario 4 – False Alarm Rate = 0.1

Action	Number of Calls	Total Time	Percentage of Total Time	Average Time
Execution of detector	10	5.150 s	90.7 %	0.515 s
Reading image file	10	0.491 s	8.6 %	0.048 s
All other actions	-	0.040 s	0.7 %	0.004 s
Total	-	5.681 s	100 %	0.568

The aforementioned experiments demonstrated the potential use of the proposed cascade object detector for hardhat identification. The algorithm is robust against variations of orientations, colors, and lighting conditions. The performance of the classifiers drops when the contrast against the background is low, and the system becomes more susceptible to missing positive instances. On the other hand, although images with a lower resolution require a significantly lower computational power, hardhats cannot be detected if their size becomes smaller than a certain threshold, hence the importance of training a detector according to a specific image resolution.

2.7. Conclusion and Future Work

Despite numerous measures and code requirements aiming at improving construction safety and health management, construction is still considered one of the most dangerous industries in the world. Workers are exposed to hazards on a daily basis, and could not be fully trusted to abide by safety regulations. Safety officers are therefore employed to monitor sites and detect any non-compliance with safety regulations. In order to automate and improve the safety inspection process, research efforts have been resorting to information technology and computer vision techniques. This paper assessed the effectiveness of existing computer vision algorithms, in particular widely adopted object detection/recognition methods, in automatically detecting hardhat-wearing on construction sites. Experiments were conducted and results highlighted that a well-trained cascade classifier was found to be robust under various scenarios and conditions. Additionally, it was proven to be relatively time-efficient and in a real-time application, it is capable of scanning for violations every 2 seconds. The process can even be expedited by reducing the resolution of the training and test images.

Further testing is required to evaluate the accuracy of different object detection algorithms, as well as to explore the potential use of heat cameras besides digital imagery. Future studies will aim as well to integrate this detection into a complete framework that: (1) detects first mobile workers, (2) scans for hardhats in the upper part of the detected worker region to speed up the detection process, and (3) issues a safety alarm or warning when a safety violation is detected. Further work will also look into improving the accuracy of the hardhat detection and eliminating false detections by combining the cascade classifier with image and color segmentation techniques.

References

Ahonen T., Hadid A., and Pietikäinen M. (2006). Face Description with Local Binary Patterns: Application to Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Aksoy M. S., Torkul O., and Cedimoglu I. H. (2004). An industrial visual inspection system that uses inductive learning. *Journal of Intelligent Manufacturing, Expanded Academic.*

Bay H., A. Ess, T. Tuytelaars, and L. Van Gool (2008). SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding*, 110 (3), 346–359.

Brilakis I., Park M.W., and Jog G. (2011). Automated vision tracking of project related entities. *Advanced Engineering Informatics*. 25 (4) 713-724.

Brunelli R. (2009). Template Matching Techniques in Computer Vision: Theory and Practice. *Wiley*.

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).

Cheng T., Venugopal M., Teizer J., Vela P. A. (2011). Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments. *Automation in Construction*, 20(8), 1173–1184.

Chi S., Caldas C. H., and 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, 199-211.

Cyganek B. (2013). Object detection and recognition in digital images: theory and practice.

Dalal N. and Triggs B. (2005). Histogram of Oriented Gradients for Human Detection. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Dickscheid T., Schindler F., and Förstner W. (2011). Coding Images with Local Features. *International Journal of Computer Vision*.

Dimitrov A. and 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*, 28 (1), 37 – 49.

Ding L., Yu H. L., Zhou C., Wu X., and Yu M. H. (2010). Safety risk identification system for metro construction on the basis of construction drawings. *Automation in Construction*, 27, 120-137.

Du S., Shehata M. and Badawy W. (2011). Hardhat Detection in Video Sequences based on Face Features, Motion and Color Information. *3rd International Conference on Computer Research and Development*.

Gheisari M., Irizarry J., and Walker B. N. (2014). UAS4SAFETY: The Potential of Unmanned Aerial Systems for Construction Safety Applications. *Construction Research Congress*.

Hamledari H., McCabe B., and Davari S. (2017). Automated computer vision-based detection of components of under construction indoor partition. *Automation in construction*, 74, 78 – 94.

Health and Safety Executive (2014). Health and safety in construction sector in Great Britain. Available at: http://www.hse.gov.uk/statistics/industry/construction/construction.pdf

Hwang S. (2012). Ultra-wide band technology experiments for real-time prevention of tower crane collisions. *Automation in Construction*, 22, 545–553.

Kasirossafar M. and Shahbodaghlou F. (2013). Application of visualization technologies to design for safety concept. *Gateway to a Better Tomorrow – Proceedings of the 6th Congress on Forensic Engineering*, San Francisco.

Khoury H., Chdid D., Oueis R., Elhajj I, and Asmar D. (2015). Infrastructureless approach for ubiquitous user location tracking in construction environments. *Automation in Construction*, 56, 47-66.

Kim Y. S., Lee J. H., Yoo H. S., Lee J. B. and Jung U. S. (2009). A performance evaluation of a Stewart platform based Hume concrete pipe manipulator. *Automation in Construction*, 18 (5), 665-676.

Kyriacou T., Bugmann G., and Lauria S. (2005). Vision-based urban navigation procedures for verbally instructed robots. *Robotics and Autonomous Systems*.

Lee G., Cho J., Ham S., Lee T., Yun S. H., and Yang H. J. (2012). A BIM- and sensor-based tower crane navigation system for blind lifts. *Automation in Construction*, 26, 1 – 10.

Leutenegger S., Chli M., and R. Siegwart (2011). BRISK: Binary Robust Invariant Scalable Keypoints. *Proceedings of the IEEE International Conference*.

Lienhart R. and Maydt J. (2002). An extended set of Haar-like features for rapid object detection. *Proc. of IEEE International Conference of Image Processing,* 1, 900-903.

Liu C., Sharan L., Adelson E.H., and Rosenholtz R. (2010). Exploring features in a bayesian framework for material recognition. *IEEE Conference on Computer Vision and Pattern Recognition,* 239–246.

Marks E., Teizer J. (2013). Evaluation of the position and orientation of (Semi-) passive RFID tags for the potential application in ground worker proximity detection and alert devices in safer construction equipment operation. *ASCE International Workshop on Computing in Civil Engineering*, 645–652.

MathWorks, Object recognition methods in computer vision, Available at: https://www.mathworks.com/discovery/object-recognition.html

MathWorks, Object recognition methods in computer vision, Available at: https://www.mathworks.com/discovery/object-recognition.html

Memarzadeh M., Golparvar-Fard M., and 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. and Tuytelaars T. (2009). Local Image Features. *Encyclo-pedia of Biometrics,* Springer US, 939-943.

Park M.W., Elsafty N., and Zhu Z. (2015). Hardhat-Wearing Detection for Enhancing On-Site Safety of Construction Workers. *Journal of Construction Engineering and Management*.

Park M.W., Koch C., and 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.

Park M.W., Palinginis E., and Brilakis I. (2012). Detection of construction workers in video frames for automatic initialization of vision trackers. *Construction Research Congress*. 940 – 949.

Rosten, E. and T. Drummond (2006). Machine Learning for High-Speed Corner Detection. *9th European Conference on Computer Vision*, 1.

Rubaiyat A.H.M., 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. *IEEE/WIC/ACM International Conference on Web Intelligence Workshops*, 135 – 142.

Seo J., Han S., Lee S. and Kim H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, 29, 239-251.

Shrestha K., Sherestha P. P., Bajracharya D., and Yfantis E. A. (2015). Hard-Hat Detection for Construction Safety Visualization. *Journal of Construction Engineering*.

Skibniewski, M. J. (2014). Information technology applications in construction safety assurance. *Journal of civil engineering and management*, 20 (6), 788-794.

Study: Construction workers experience higher rates of injury, premature death. *Safety and Health Magazine.* Available at: http://www.safetyandhealthmagazine.com/articles/construction-workers-experience-higher-rates-of-injury-premature-death-study-2

Teizer J., Castro-Lacouture D. (2007). Combined ultrawideband positioning and range imaging sensing for productivity and safety monitoring in building construction. *Congress on Computing in Civil Engineering*, 681–688.

United States Department of Labor, Occupational safety & Health administration. Common Statistics. Available at: https://www.osha.gov/oshstats/commonstats.html

United States Department of Labor, Occupational safety & Health administration (2006). Module 13—Personal protective equipment. Available at: https://www.osha.gov/dte/

Viola P. and Jones M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*. 57 (2), 137-154.

Yan J., Park, M.W., Vela P.A., and Golparvar-Fard M. (2015). Construction performance monitoring via still images, time-lapse photos, and video streams: now, tomorrow, and the future. *Advanced Engineering Informatics*, 29 (2) 211-224.

Yoshida T. and Chae S. (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. *IEEE International Conference on Multimedia and Expo*.

CHAPTER 3

AN INTELLIGENT VISION-BASED FRAMEWORK FOR HARDHAT- WEARING DETECTION IN INDOOR CONSTRUCTION SAFETY APPLICATIONS

Abstract

The construction industry is still among the riskiest industries in the world despite numerous enforced safety regulations on jobsites. Workers are subject to getting injured from falls, slips, trips, or getting struck by falling objects. Hence, safety programs have been according great emphasis on proper use of personal protective equipment (PPE) by deploying safety officers on construction sitesto ensure proper implementation of those requirements. However, thecurrent practice of supervising large construction areas could prove to be manual, tedious, and ineffective. Therefore, this study aims at creating a comprehensive framework that can automatically and efficiently detect any non-compliance with safety rules and regulations, in particular a failure to wear a hardhat, using computer vision techniques applied on videos captured from indoor construction sites. This is achieved by: (1) isolating mobile workers from the captured scene by means of a novel motion detection algorithm, and (2) detectinghardhat in the identified region of interest using object detection together with image and color segmentation. Several experiments were conducted and results highlighted that the newly developed motion detection algorithm showed an improved accuracy compared to common background subtraction methods, and the hardhat detection algorithm achieved high precision and recall.

Keywords

Construction safety, PPE, hardhat-wearing, motion detection, object detector, color segmentation.

3.1. Introduction

Construction workers are exposed to a variety of hazards on a daily basis. Risks include falls, slips, trips, getting struck by falling or flying objects, electrocutions, being caught in-between equipment or objects, and being crushed by a collapsing structure (OSHA 2015). And in spite of numerous measures and provisions established to create a safer working environment, construction still produces one of the highest injury and fatality rates between all sectors. 937 construction workers were fatally injured in the United States in 2015, which constitutes 19.3% of the total workplace deaths, and a 2 percent rise from the previous year and the highest number of casualties in the construction field since the 975 reported cases in 2008 (Bureau of Labor Statistics 2015). In Canada, construction had the highest number of death among all industries and accounted for 23.3% of the total work related fatal injuries between 2008 and 2010 (CBC News 2012). In the United Kingdom, the rate of fatal injury per worker for the construction industry is over 3.5 times greater than the average rate across all other sectors (Health and Safety Executive 2015). In order to avoid the legal, financial, and moral consequences implied by the tragic loss of life on a construction site, safety codes and regulations have been developed. Those include the “Safety and Health Regulations for Construction” in the United States, the “Occupational Health and Safety Regulations” in Canada, and the “Health and Safety in Construction” in the United Kingdom. The Construction Design and Management Regulations (2007) assign the responsibility of planning, coordinating and managing health and safety through the entire project to clients, supervisors, designers and contractors alike (Ashworth 2013).

Remarkable efforts have been realized to improve construction safety management and the implementation of safety regulations is vastly prioritized by owners, contractors and legal authorities in the majority of developed countries (Sherrat et al. 2013). On the other hand, most developing countries display a weak commitment to safety requirements and a poor safety performance (Teo et al. 2008, and Chiocha et al. 2011). In a recent study about the state of construction safety in Lebanon, a middle eastern developing country, contractors stated that the absence of a safety and health management system on their construction sites was caused by the lack of law enforcement, the high cost and time of safety implementation (Awwad et al. 2015).

Generally, safety officers are employed to monitor job sites and ensure proper compliance with safety regulations. They educate workers about potential hazards through weekly safety trainings and toolboxes, and enforce the appropriate use of personal protective equipment (PPE) according to the specified task, including hardhats, protective eyewear, high visibility vests, gloves, safety boots, fall protection and other means where necessary. Safety equipment and more precisely hardhats are a major component in any safety program and have proven capabilities in protecting against head injuries resulting from impact with falling or flying objects. In fact, 84% of the total head injuries in the construction industry are sustained by workers not wearing hardhats (OSHA 2006). However, the task of actively supervising a large number of workers on a sizeable project, using manual traditional methods, could prove to be inefficient, tedious, and requires an abundance of human and financial resources. For this reason, optimizing and automating the current construction safety process would offer more incentive to clients and contractors to implement the required codes and regulations.

Therefore, this paper presents a novel safety inspection framework capable of efficiently identifying individuals not wearing hardhats by applying a series of computer vision techniques and algorithms on photos and videos captured from indoor jobsites. The overall system consists of two major components. In the first part, a motion detection algorithm is developed. The algorithm identifies moving objects in a set of consecutive images using standard deviation differences between the images. A standard classifier is then applied to the active region to identify the presence of any individual. In the second part, a cascade object detector is utilized to search for a hardhat in the top part of the identified personnel and image and color segmentation techniques are then used to eliminate false detections.

3.2. Background

Traditional safety methods and programs have been the subject of numerous studies. Construction safety research includes behavior-based safety (BBS) management (Lingard and Rowlinson 1997), integration of safety and construction schedules (Chen et al. 2000, and Coble et al. 2000), development of safety training systems (Aranda 2000), analysis of accident causes (Hinze et al. 1998, and Hinze et al. 2005), and safety hazard identification (Carter and Smith 2006). Recently, the trend has shifted to computer-based and IT systems to automate the current safety practices. For instance, Skibniewski (2014) reported that 136 articles focusing on IT applications in this field were published between 2006 and 2014. Research topics included sensor-based systems, robotics and manipulators, and information analysis, management and reporting. Sensor-based systems in the field of construction safety rely on a wide range of "Real Time Locating Systems" (RTLS) that include "Radio Frequency Identification" (RFID) sensors, "Ultra-wide Band" (UWB) sensors, and laser sensors, to create early warning systems that prevent collision accidents due to blind spots (Teizer and Castro-Lacouture 2007, Yoshida and Chae 2010, Hwang 2012, Lee et al. 2012, Marks and Teizer 2013, and Cheng and Teizer 2014). On the other hand, robotics and manipulators have been designed to replace human workforce when encountered with tasks performed under dangerous conditions. Kim et al. (2009) evaluated the performance of the Hume Concrete Pipe Manipulator (HCPM), a system commissioned by the Korea Ministry of Construction and Transport to perform pipe laying work. This task previously required laborers to work in trenches and was considered one of the most dangerous construction operations. In addition to eliminating potential safety hazards, analysis of the HCPM's performance indicated a 65% improvement in productivity and a 33% reduction is costs. Finally, information analysis, management and reporting includes the use of BIM for safety planning and schedule-workspace interference visualization, to secure work performance safety and decrease collision risk between resources Ding et al. (2012).

Among other IT systems adopted in automating construction processes, in particular safety inspections, computer vision proved to be of paramount importance. In fact, computer vision techniques have been used to automate field observations and replace existing time-consuming data collection methods. Applications include progress monitoring, productivity analysis, object detection, tracking and motion detection, and action recognition (Yang et al. 2015). Texture recognition systems were used to identify on-site material for construction progress monitoring (Liu et al. 2010, and Dimitrov and Golparvar-Fard 2014). Hamledari et al. (2016) presented a computer vision based system that relies on shape and color analysis, to detect indoor partition components such as studs, insulation, electrical outlets and drywall sheets. Based on detection results from images and videos from construction sites, the system classifies the current stage of work into one of five possible stages: (1) framing, (2) insulation, (3) installed drywall, (4) plastered drywall, and (5) painted partition. The obtained information may then be utilized to update corresponding BIM models and schedules.

Computer vision-based systems have also found a direct application in construction safety through identifying unsafe acts, hazardous situations and awkward postures (Seo et al. 2015). Some systems have been typically collecting data from multiple static cameras to identify, track and calculate 3D coordinates of construction entities in each frame and alarming personnel of unsafe conditions (Brilakis et al. 2011, and Park et al. 2012a). Chi et al. (2009) proposed an algorithm that identifies and tracks objects on construction sites using a high-frame-rate range sensor. Image matching algorithms were used to classify objects using a model database and through comparison with previous scans. Memarzadeh et al. (2013) presented a computer vision-based framework to automatically detect construction workers and equipment using videos streamed from construction sites. They utilized an algorithm based on Histogram of Oriented Gradients and Colors (HOG + C) to make the semi-automated detection less time-consuming. On the other hand, motion detection systems were used in construction safety applications and generally rely on background subtraction algorithms to detect foreground and moving objects (Park et al. (2015). More specifically, Park et al. (2012b) utilized standard background subtraction to detect moving blobs in an image. Within the moving foreground, HOG features are used to identify people, then color gradients are used to distinguish between pedestrians and construction workers, based on the color of the safety vest. Furthermore, other research has been recently carried out to facilitate, improve, and automate appropriate PPE usage enforcement procedures (Bajracharya 2013). 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 was detected using existing face detection algorithms based on Haar-Like features, introduced by Lienhart and Maydt (2002) and Viola and Jones (2004). 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 not real-case scenarios from construction sites.

Furthermore, Gheisari et al. (2014) evaluated the applicability of unmanned aerial systems (UAS) or camera-equipped drones in construction safety, in particular hardhat-wearing detection, by providing safety personnel with real-time visual access to jobsites. Results revealed the importance of adopting drones in safety applications. However, a large-size interface on a tablet device is required to provide a precise view of the jobsite. Additionally, the detection process is manual and required the safety officer's presence and perception. Shrestha et al. (2015) proposed an algorithm that detects workers using standard face detection then applies the edge detection technique on the region directly above the worker's head. The system detects a hard-hat if its outline is a semicircle and its color is red. However, their program 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. No validation of the efficiency of their system in an actual construction environment was also realized. Park et al. (2015) detected hard-hats using a support vector machine (SVM) classifier, as part of a complete framework that aims at enhancing on-site safety conditions. The algorithm is based on shape recognition and utilizes histogram of oriented gradient (HOG) features to describe the cap style shape of the hard-hat. While the proposed framework was capable of detecting a hardhat under various 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. A recent effort by Rubaiyat et al. (2016) utilized Circle Hough Transform (CHT) feature extraction that can identify the circular shape of the hardhat coupled with color analysis using threshold-based color segmentation in RGB space to create an effective helmet detection system. Their method was capable of detecting hardhats at a rate of 79.1%.

The algorithm identifies all yellow, blue, red and white circles as hardhats. Nevertheless, the effect of brightness on color values was not addressed.

Besides the aforementioned limitations of existing studies, prior algorithms for automated hardhat detection achieved moderate rates of precision and recall. An object detector based on HOG features proved promising, but could not achieve high rates of precision and recall simultaneously. Additionally, the detection of a region of interest for hardhat identification, defined as the upper part of an individual, was never investigated. Therefore, this study aims to create a comprehensive framework that first detects individuals and defines the region of interest (i.e. hardhat area), then uses an object detector coupled with color analysis and segmentation techniques to accurately identify the presence of worn hardhats. Comparing the number of detected personnel against the number of detected hardhats would allow for the recognition of any non-compliance with the required safety rules and regulations.

3.3. Methodology

In order to achieve the aforementioned objectives, this paper presents a novel framework that consists of four major stages (Figure 3.1): (1) Data Acquisition, (2) Motion detection, (3) Human detection in moving region, and (4) Hardhat detection in top human region. This paper mainly focuses on the second and fourth components of the proposed framework that were implemented using Matlab 2016a.

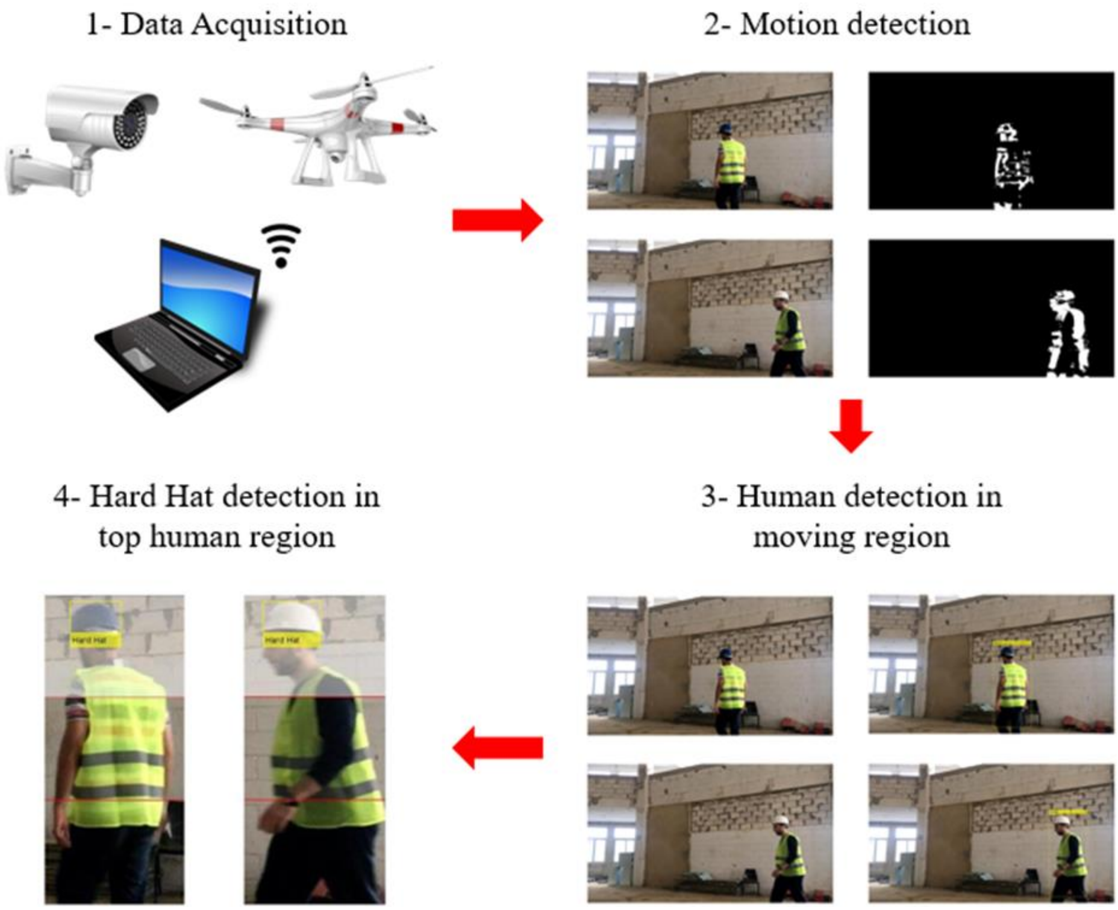

Figure 3.1: The four stages of the hardhat detection framework

The first component, data acquisition, involves the collection of images and videos from construction sites. Surveillance and CCTV cameras have become vastly employed on construction sites to monitor workers, evaluate the site and determine the resource usage. They offer a cheap and effective data acquisition method but require and installation process which can pose a disadvantage in a dynamic construction environment. Camera-equipped drones and Unmanned Aerial Systems (UAS) are infrastructure-less, and can be programmed to fly autonomously to the required destinations. The aim of this study, however, is to create an automated hardhat detection method independentof what technology was used to capture images and videos of construction sites.

In the second part of the proposed framework, motion detection utilizes a set of computer vision techniques to identify moving objects in videos and eliminates the need for complex and expensive motion sensors. In this context, this study presents a novel motion detection technique based on the pixel to pixel standard deviation differences between a set of consecutive images. To demonstrate the effectiveness of this method, the study first presents a classic background subtraction algorithms, then compares its accuracy against that of the developed algorithm, using a set of videos captured from construction sites.

In the the third framework component, detected moving regions can be filtered to identify the presence of a person, instead of the whole image. Human detection generally relies on various object classifiers widely known in the literature and previously discussed in section 2. In our study, it was assumed that all moving objects larger than a minimum required size are identified as humans. This assumption is sound for indoor construction applications as every movement is either performed or caused by human intervention. For outdoor construction applications, possible cases of wrong identification include tower crane operations and objects displaced by wind. More specifically, mistankenly identifying an object as a construction worker followed by the non-identification of a hardhat in the corresponding region can result in a false detection and accordinglyconstitute a potential safety hazard that requires intervention.

In the final stage of the framework, the algorithm searches for a hardhat in the top third of the region identified as an individual. This operation greatly reduces the duration of the required computation and eliminates the possibility of detecting hardhats placed on the ground or not worn by construction personnel. The detection process utilizes a cascade object detector and image and color segmentation to eliminate false detections.

In the following subsections, the second and fourth components are thoroughly explained.

3.3.1. Motion detection using computer vision

The most widely used approach to detect moving components in videos captured from a static camera is known as background substraction (Piccardi 2004). The aim of the algorithm is to separate the scene into a “Foreground” comprised of moving objects, and a still “Background”. The method identifies moving objects based on differences between the current frame and a reference frame. The basic form of background subtraction involves the calculation of the absolute difference between the two images, the current frame and the reference frameCalculation is realized on a pixel to pixel level and can be performed in any color space, including, but not limited to, RGB, CIE LAB, and grayscale . RGB is a classical representation of colors in which Red, Green, and Blue are combined in different ways to create an array of colors. An image with a height of 720 pixels and a width of 1280 pixels is stored in the RGB color space as a 720 × 1280 × 3 three dimensional array of three matrices, the Red, Green, and Blue matrices respectively. The CIE LAB (Hunter 1948) color space is another three dimentional representation where a given color is represented by the combination of three matrices: (L) lightness component, (a) Green – Red component, and (b) Blue – Yellow component. On the other hand, the grayscale is a one dimensional representation of a pixel based on the value of its intensity. A grayscale image is entirely composed of shades of gray. The intensity value can be calculated based on the

RGB values according to Eq. 3.1 and the background subtraction happens according to Figure 3.2 while using Eq 3.2.

$$Intensity = [0.299 \quad 0.587 \quad 0.114]\begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.1}$$

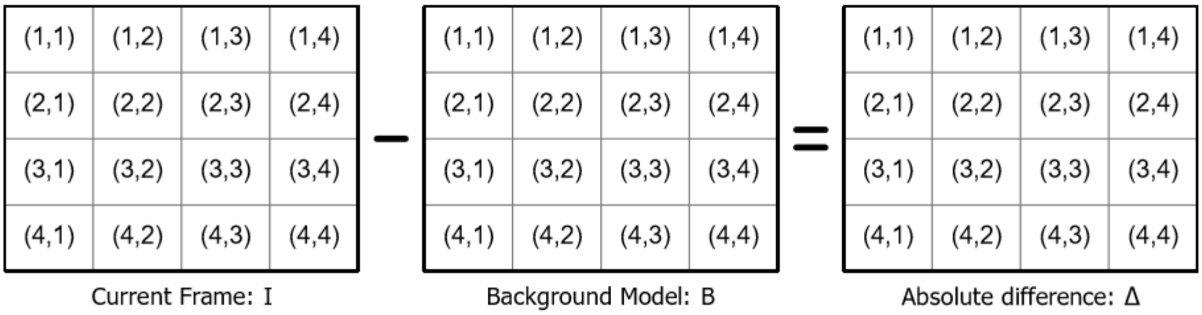

Figure 3.2: Simple Background Subtraction in Grayscale color space

$$\Delta_{(i,j)} = |\, I_{(i,j)} - B_{(i,j)} \,| \tag{3.2}$$

Every pixel in the delta matrix is then classified as foreground or background by comparing it with a previously defined threshold. The calculation process is programmed in the RGB color space, taking a threshold equal to 10 for example, using the following code:

```
Background = imread('Background.jpg');
Image = imread('Current.jpg');
Delta = abs(Image - Background);
BinaryImage = Delta(:,:,1)>10 | Delta(:,:,2)>10 |
Delta(:,:,3)>10;
```

The resulting binary image has the same size as the original images whereby true values (1) represent moving pixels and false values (0) represent still pixels. Using the CIE LAB color space provides the advantage of ignoring the differences in the values of lightness when creating the binary image as shown in the code snippet below::

```
Background = rgb2lab(imread('Background.jpg'));
Image = rgb2lab(imread('Current.jpg'));
Delta = abs(Image - Background);
BinaryImage = Delta(:,:,2)>10 | Delta(:,:,3)>10;
```

The modeling of the background image has been the subject of many studies. Methods included running Gaussian average, temporal median filter, Gaussian mixture models, Kernel Density Estimation, sequential Kernel Density approximation, co-occurrence of image variations and Eigenbackgrounds (Wren et al. 1997, Lo and Velastin 2001, Stauffer and Grimson 1999, Elgammal et al. 2000, Han et al. 2004, Seki et al. 2003, Oliver et al. 2000). Methods vary in terms of calculation speed and accuracy and are only applicable for videos captured using a perfectly static camera (Piccardi 2004). In matlab, the “foregroundDetector” system object performs background subtraction operations and computes the foreground mask using Gaussian Mixture Models (GMM) (MathWorks 2017). The object can be programmed to detect moving components in videos, as shown in the code snippet below:

```
Detector = vision.ForegroundDetector;
```

The created detector can then be called upon using the “step” function, and given a series of video frames, it performs the required computations and returns the foreground mask, as a binary image similar to the one created earlier. Implementation of the Matlab built-in background subtraction algorithm is realized using the following code:

```
videoSource = vision.VideoFileReader('Video.mp4');
while ~isDone(videoSource)
      frame = step(videoSource);
      BinaryImage = step(detector, frame);
End
```

On the contrary, the proposed motion detection algorithmdoes not involve any sort of background subtraction. Instead, “Foreground” or moving regions are identified based on the standard deviations between the values in a particular slot (pixel position) in a set of consecutive images. Therefore, the classification is based on the values of the “Standard Deviation Matrix” (SDM) instead of the Delta Matrix. In order to calculate the SDM, the system first needs to acquire a set of (n) consecutive images. Our implementation of the algorithm is based on the RGB color space (Figure 3.3) using the following code:

```
In = imread('Image.jpg');
Rn = In(:,:,1)
Gn = In(:,:,2)
Bn = In(:,:,3)
```

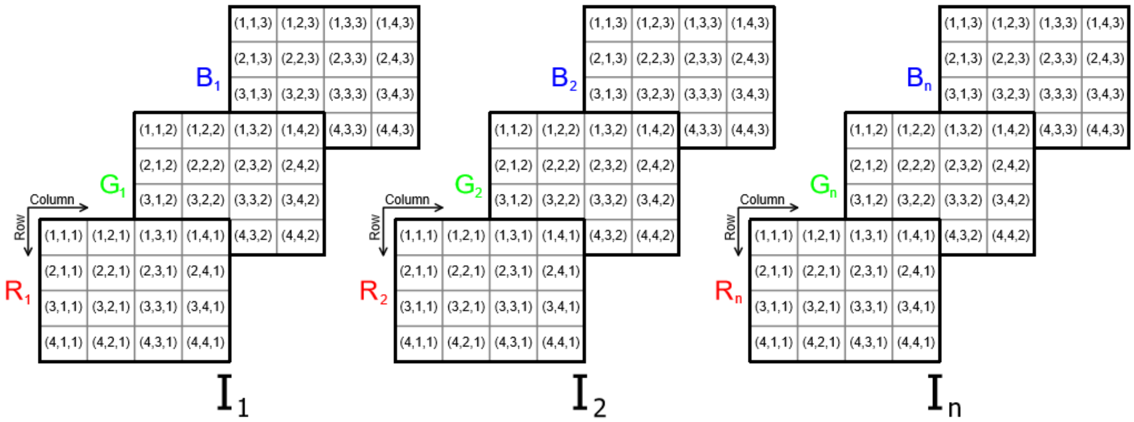

Figure 3.3: Consecutive images stored in the RGB color space

The calculation of the values in the SDM is illustrated in Figure 3.4, and is based on the following general formula:

$$\mathrm{SDM}_{(i,j,k)} = \text{Standard Deviation of } [\, I_{1\,(i,j,k)}, I_{2\,(i,j,k)}, \dots, I_{n\,(i,j,k)} \,] \tag{3.3}$$

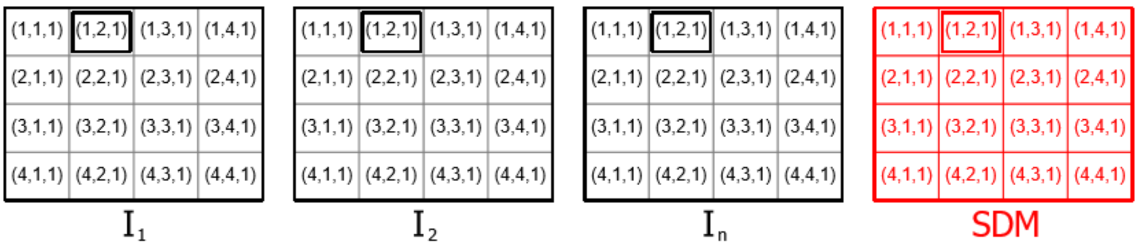

Figure 3.4: Calculation of the Standard Deviation Matrix

Most importantly, the calculation of standard deviation values in-between matrices is not commonly used and does not have an implemented Matlab function. For this reason, a series of transformations and operations are required. The first step is called linearization, whereby an a by b matrix is transformed into a 1 by a × b linear array (Figure 3.5) as shown in the code snippet below:

```
a = size(r1,1); b = size(r1,2);
LRn = reshape(Rn,1,a*b);
LGn = reshape(Gn,1,a*b);
LBn = reshape(Bn,1,a*b);
```

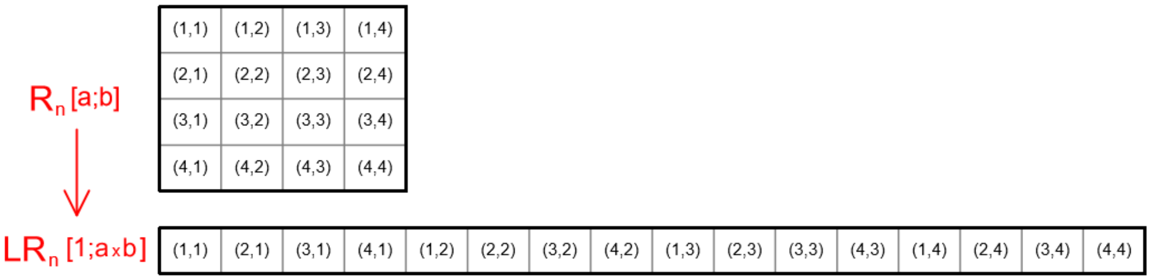

Figure 3.5: Example of Linearization – R matrix

The next step involves the creation of three general matrices, GR, GG, and GB, having a size of n by a × b, and corresponding to the Red, Green, and Blue matrices respectively. Each row of the general matrix is extracted from the corresponding linearized matrix previously calculated. Each column represents all pixel values in the input images for a specific position. Once a general matrix is created, the “std” function is used to calculate the standard deviation of the values in each column, thus obtaining a linearized form of the standard deviation matrix (SLR, SLG, or SLB). The previous operation is illustrated in Figure 3.6 for the GR matrix, and is implemented, for n equal to 5, using the following code snippet:

```
GR = [LR1;LR2;LR3;LR4;LR5];
SLR = std(GR);
```

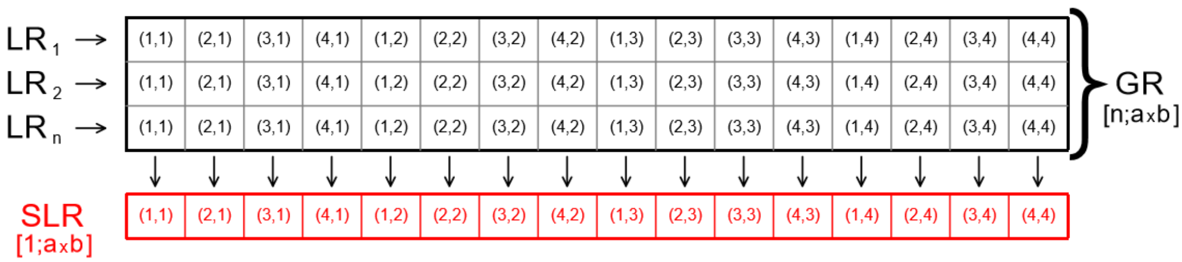

Figure 3.6: Calculation of Linear Standard Deviation Matrix

Once the three linear standard deviation matrices are calculated, the final step involves reshaping these matrices into the original a by b dimensions of the input images (Figure 3.7), to obtain the three Standard Deviation Matrices, using the “reshape” Matlab function:

```
SR = reshape(SLR,a,b);
SG = reshape(SLG,a,b);
SB = reshape(SLB,a,b);
```

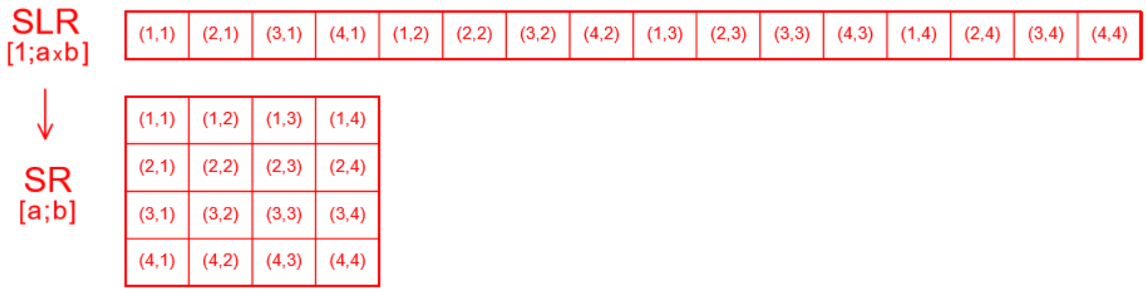

Figure 3.7: Reshaping the Linear Standard Deviation Matrix into the Standard Deviation Matrix

The binary image can be finally created using thresholding methods similar to the ones used for background subtraction. In this context, foreground or moving pixels are identified for having a calculated standard deviation greater than five times the average standard deviation for the corresponding matrix, according to the following code:

```
SdR = mean(SLR);
SdG = mean(SLG);
SdB = mean(SLB);
BinaryImage = SR > 5*SdR & SG > 5*SdG & SB > 5*SdB;
```

Similar to background subtraction, the resulting binary image has the same dimensions as the input images. In both methods, the size and location of the moving components are obtained from the binary image by applying blob analysis and morphological operations. Blob analysis is an algorithm that calculates the properties of connected pixels (Blobs) in a binary image. These properties include the area, centroid, bounding box, and blob count. The bounding box output of the blob analysis is a 1 by 4 array that contains information about the location of the detected element. The four parameters used to describe the location are: (1) x, (2) y, (3) width, and (4) height (Figure 3.8).

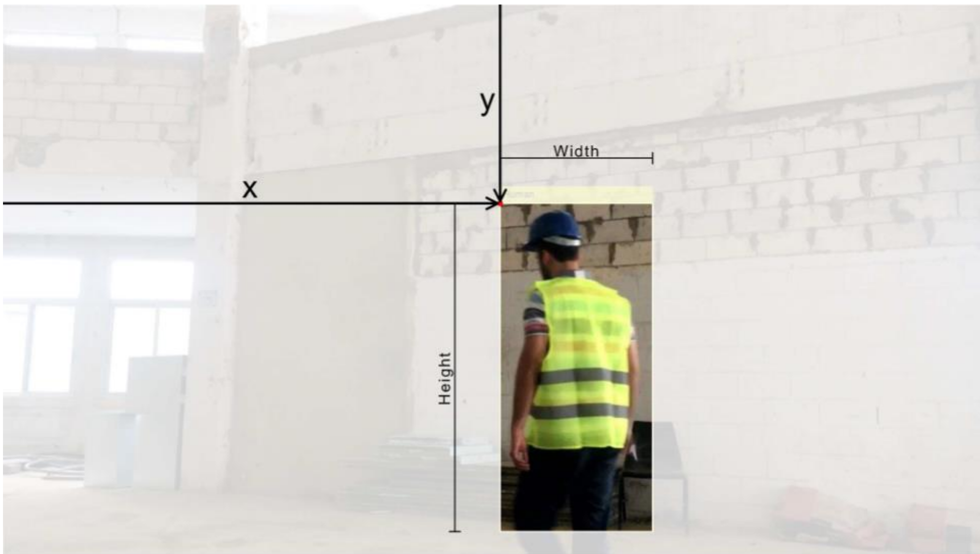

Figure 3.8: Bounding Box parameters of detected region

Morphological operations are image processing operations, applied prior to the blob analysis to remove noise and texture distortions in binary images, include removing isolated pixels, bridging nearby unconnected pixels then performing morphological closing (dilation followed by erosion). Once an individual is identified, the later stages of the framework involving the detection of hardhats get initiated. The proposed motion detection procedure can be summarized in Figure 3.9.

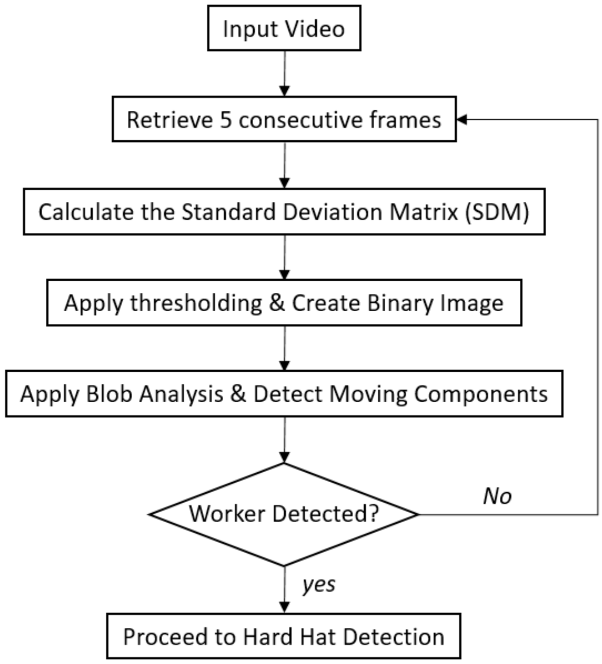

Figure 3.9: Proposed motion detection algorithm flowchart

3.3.2. Hardhat detection using computer vision

Automated hardhat detection is an integral component of the overall framework. A previous study investigating the use of computer vision techniques to detect hardhats from images and videos concluded that training an object detector was the most suitable method (Mneymneh et al. 2017). The study evaluated three computer vision techniques, including feature detection, extraction and matching, template matching, and cascade object detectors, and concluded that the latter achieved better results in terms of accuracy and computational time. For this reason, a cascade object detector based on Histogram of Oriented Gradients (HOG) features was used in this study. The training dataset is comprised of 75 positive instances and 164 negative ones. The number of stages and the acceptable false alarm rate are set to 7 and 0.1 respectively. To create the classifier, the “trainCascadeObjectDetector” function in Matlab writes a trained cascade detector XML file, which can be later used to identify and count the hardhats in an input image, as shown in the following code snippet:

```
trainCascadeObjectDetector('HOG_7_10.XML' , positiveInstances ,
negativeFolder , 'FalseAlarmRate' , 0.10 , 'NumCascadeStages' ,
7 , 'FeatureType' , 'HOG');
hatDetector = vision.CascadeObjectDetector('HOG_7_10.XML');
I = imread('TestImage.jpeg');
Detection = step(hatDetector,I);
```

The HOG features utilized to identify a hardhat are capable of describing its semi-circular shape, thereby enabling the created classifier to detect different colors and orientations. However, it also makes the detector susceptible to wrong detections since the semi-circular profile of the hardhats can also be extracted from other, irrelevant objects (Park et al 2015). To improve detection accuracy, two additional steps are applied. First, the object detector only searches for hardhats in the upper region of the identified personnel, which reduces the size of the tested sample, eliminates the possibility of detecting a hardhat on the ground and not actually worn by a worker, and greatly reduces the computational time required to scan through the entire image. The second step involves color segmentation, whereby positive detections not conforming with possible color schemes of a hardhat are filtered out.

The first step requires separating the top third of the detected region into a new image, based on its coordinates in the input frame (Figure 3.10), using the following code:

```
Region =
Image(bbox(2):bbox(2)+bbox(4)/3,bbox(1):bbox(1)+bbox(3),:);
```

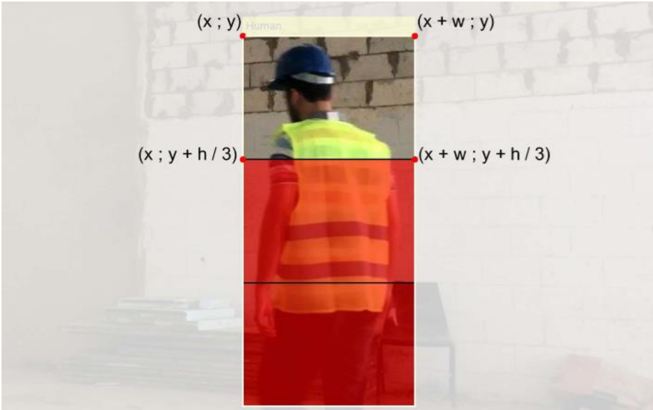

Figure 3.10: Coordinates of the upper third of the detected region

The cascade object detection can then be applied on the new image and the output is a new bounding box object describing the location of the hardhat – if detected – similar to the detected region in Figure 3.8. On the other hand, possible color patterns are required to perform the second additional step, which is color segmentation. For this reason, numerous images of blue, orange, and white hardhats are collected under different lighting situations, and cropped so that the resulting image would only contain portions of the hardhat. In Matlab, the average and standard deviation of pixel values are calculated for each image, using various color spaces to determine the most accurate representation of the hardhat color. The blue hardhat is characterized by a relatively dark appearance under usual on-site lighting conditions. For this reason, the average values of Red, Green, and Blue in the RGB representation of 102 sample images 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 (Figure 3.11) are not a definitive representation of a blue color.

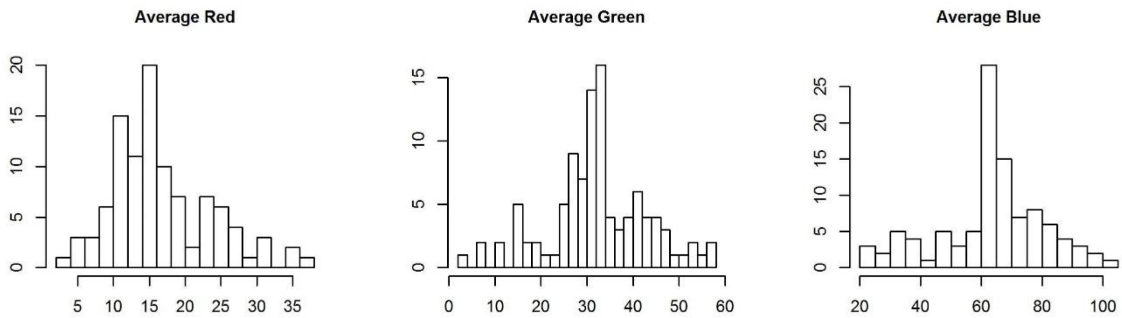

Figure 3.11: Average Red, Green, and Blue values in images of a blue hardhat

The standard deviationbetween pixel values in every image was also large compared to the average values (Figure 3.12), which also highlights the inaccuracy of the RGB representation.

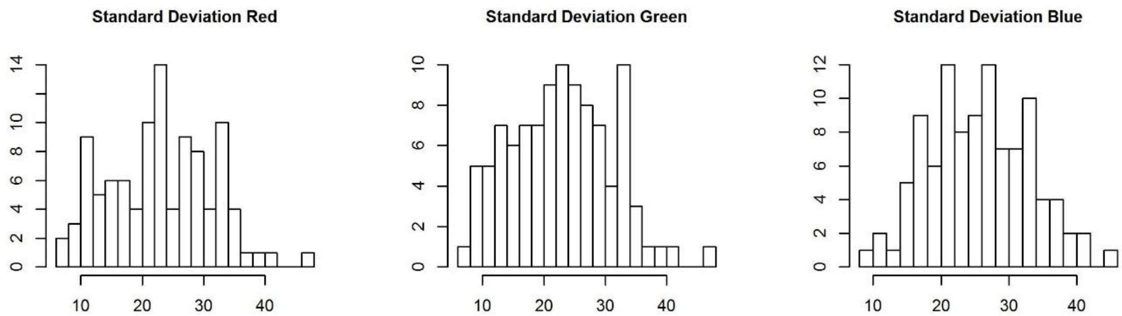

Figure 3.12: Standard deviation of Red, Green, and Blue values in images of a Blue Hardhat

Using CIE LAB color space, the average values of Lightness, A (green – Red), and B (blue – yellow), range from 1 to 23, 2.3 to 8, and -28 to -8 respectively (Figure 3.13). The low values of lightness are a result of the dark color of the hardhat, while the neutral values of A and the negative value of B represent the blue color of the hardhat. However, standard deviation values within each image were still large (Figure 3.14).

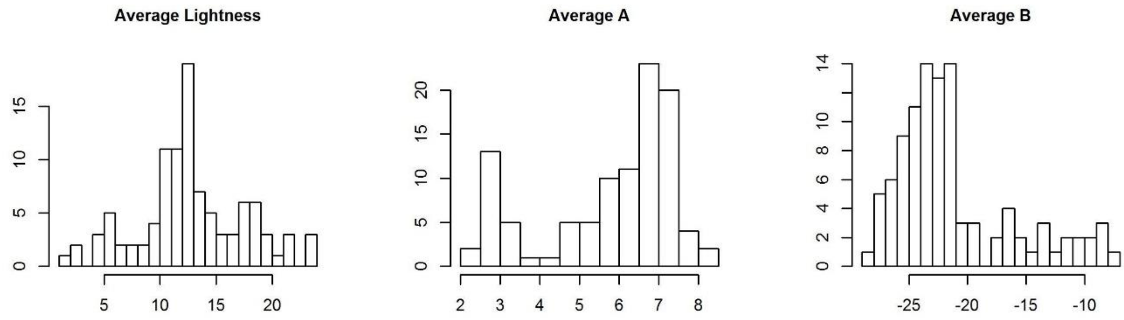

Figure 3.13: Average Lightness, A, and B values in images of a Blue Hardhat

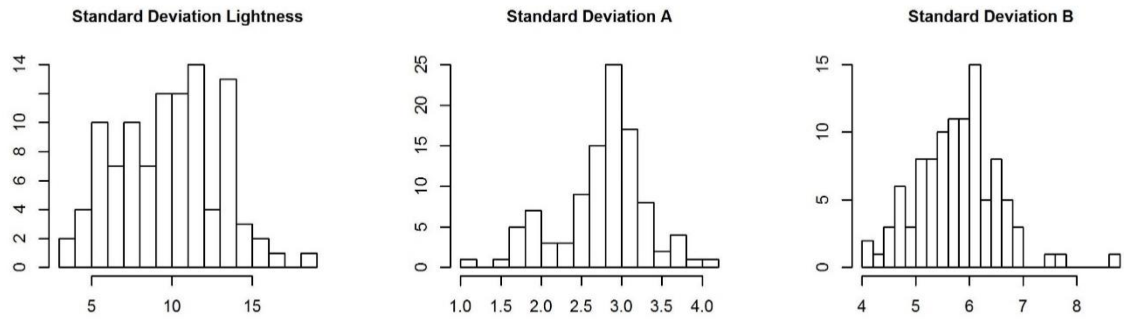

Figure 3.14: Standard deviation of Lightness, A, and B values in images of a Blue Hardhat

Another color representation that could potentially provide a more accurate definition is HSV (Joblove and Greenberg 1978). HSV stands for Hue, Saturation and Value (Brightness), and is based on how colors are conceived by the human vision. Basically, Hue refers to pure color form (Figure 3.15), Saturation refers to the amount of color, while Value refers to the brightness of the color.

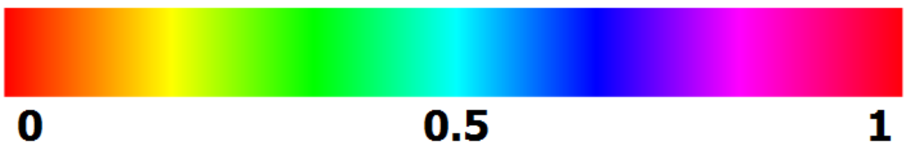

Figure 3.15: Hue scale in HSV representation

The calculated averages of Hue, Saturation, and Value for the 102 sample images range from 0.59 to 0.66, 0.66 to 0.95 and 0.08 to 0.39 respectively (Figure 3.16), providing the most accurate representation of a blue color with low brightness. The value of standard deviation for Hue were also minimal (Figure 3.17), meaning that the uniformity of the hardhat color within each image is accurately modeled.

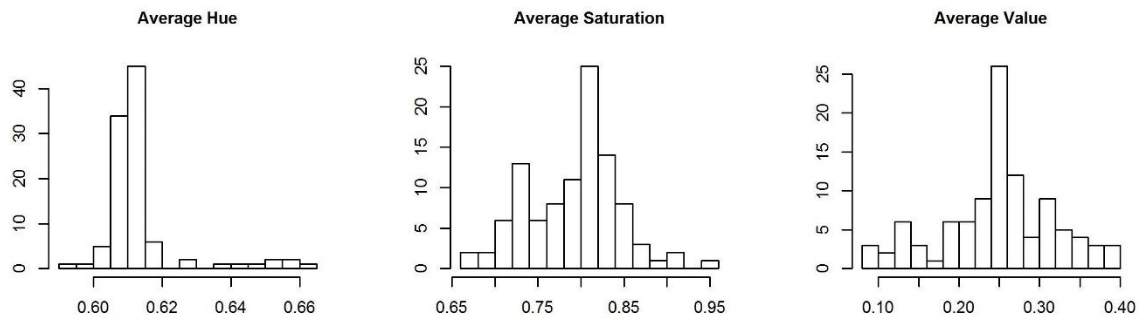

Figure 3.16: Average Hue, Saturation, and Value values in images of a Blue Hardhat

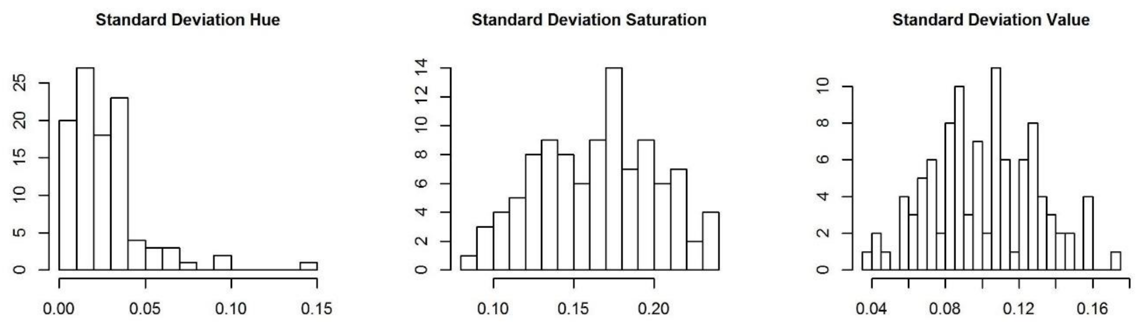

Figure 3.17: Standard deviation of Hue, Saturation, and Value values in images of a Blue Hardhat

Accordingly, a blue hardhat is defined as having an average Hue value between 0.59 and 0.66, and a standard deviation of Hue values less than 0.1. The other parameters are irrelevant. HSV color analysis for orange and white hardhats returned similar results. An orange hardhat is characterized by a Hue between 0.02 and 0.07, while a white hardhat is characterized by low values of saturation and high values of brightness. Standard deviation values were all less than 0.1.

In this study, color segmentation in HSV color space needs to be applied to the parts of images containing only the detected hardhat. The presence of another object (portion of the head, wall, etc.) would imply an incorrect calculation of the color parameters. The target area, defined as the middle third of the detected hardhat, can be obtained from the bounding box output of the cascade object detector, similar to the method used to extract the top region of the detected personnel (Figure 3.18).

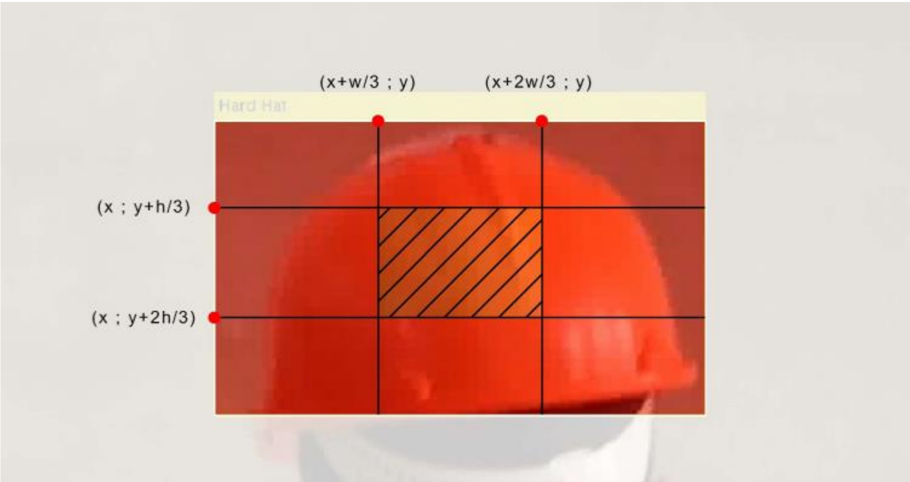

Figure 3.18: Coordinates of the hardhat target area

The algorithm retrieves the target area then calculates the required average and standard deviation values of all pixels. The object is then classified as a blue, orange, or white hardhat if the calculated parameters conform with the pre-defined color schemes, otherwise, the detection is filtered out. The hardhat detection procedure is summarized in Figure 3.19.

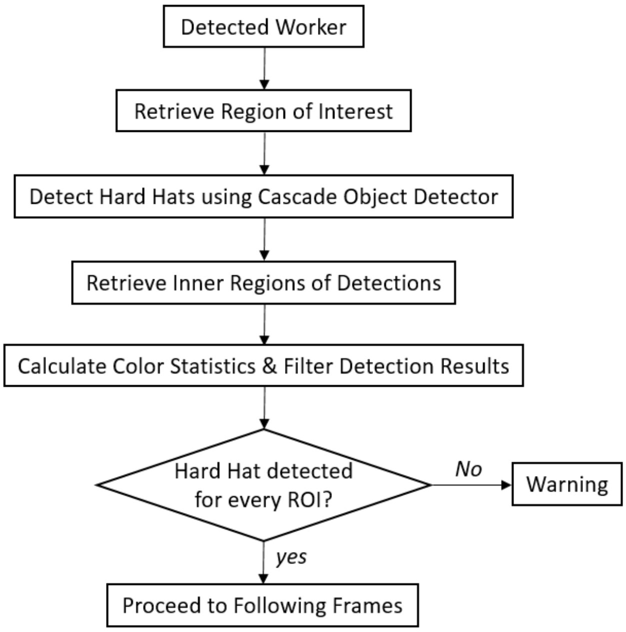

Figure 3.19: Hardhat detection flowchart

3.4. Experiments and results

3.4.1. Evaluation of the motion detection algorithm

Accurate motion detection is a critical component of the framework. Undetected moving objects cannot be identified as construction personnel and a failure to comply with required safety regulations could be missed. As such, in this section, the ability of the developed algorithm in detecting motion and pixel variation from image sequences and videos was tested. Results of the standard deviation method are then compared to those of the original background subtraction method, using the same image and video datasets.

In the first set of experiments prior to any morphological operation, a remarkable improvement of accuracy and reduction of noise was noted when using the standard deviation difference method as opposed to background subtraction using RGB and LAB color spaces. A sequence of five consecutive images is required to apply the algorithm, while only a background and an input image are required in the case of background subtraction. The sample images are extracted from a continuous video captured using a static camera. Figures 3.20, 3.21, 3.22, and 3.23 illustrate the obtained binary images using each method, prior to the application of any morphological operation.

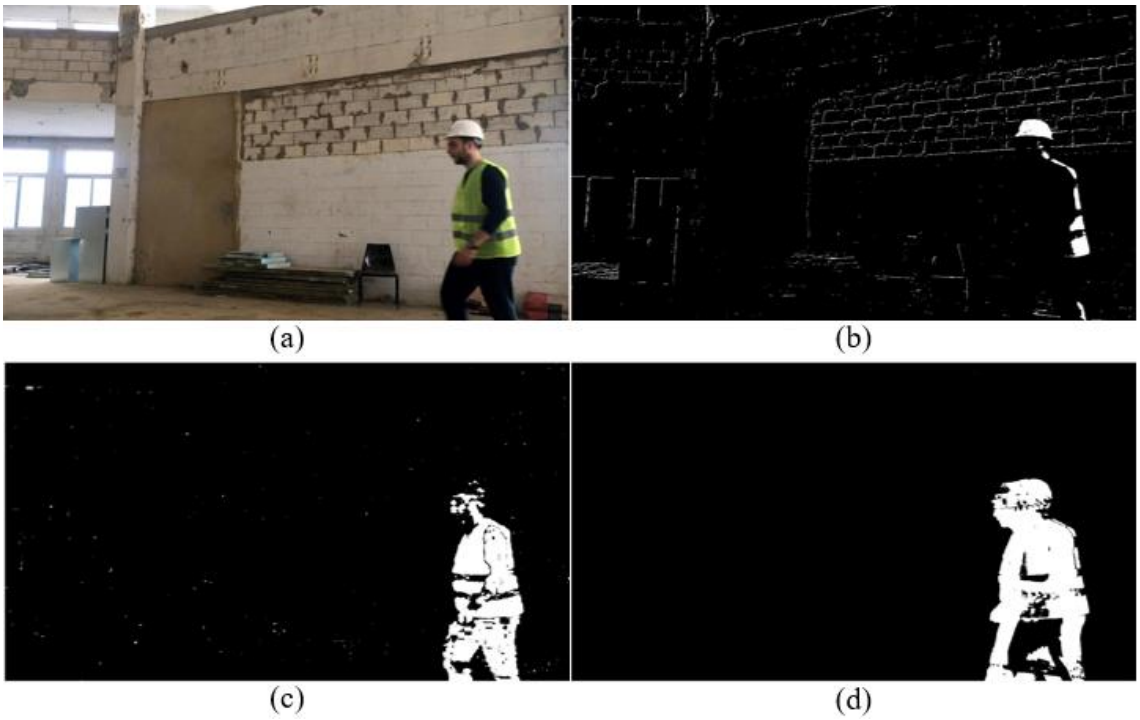

Figure 3.20: Sample image 1: (a) Original image, (b) Background subtraction in RGB, (c) Background subtraction in LAB, and (d) Standard Deviation Method

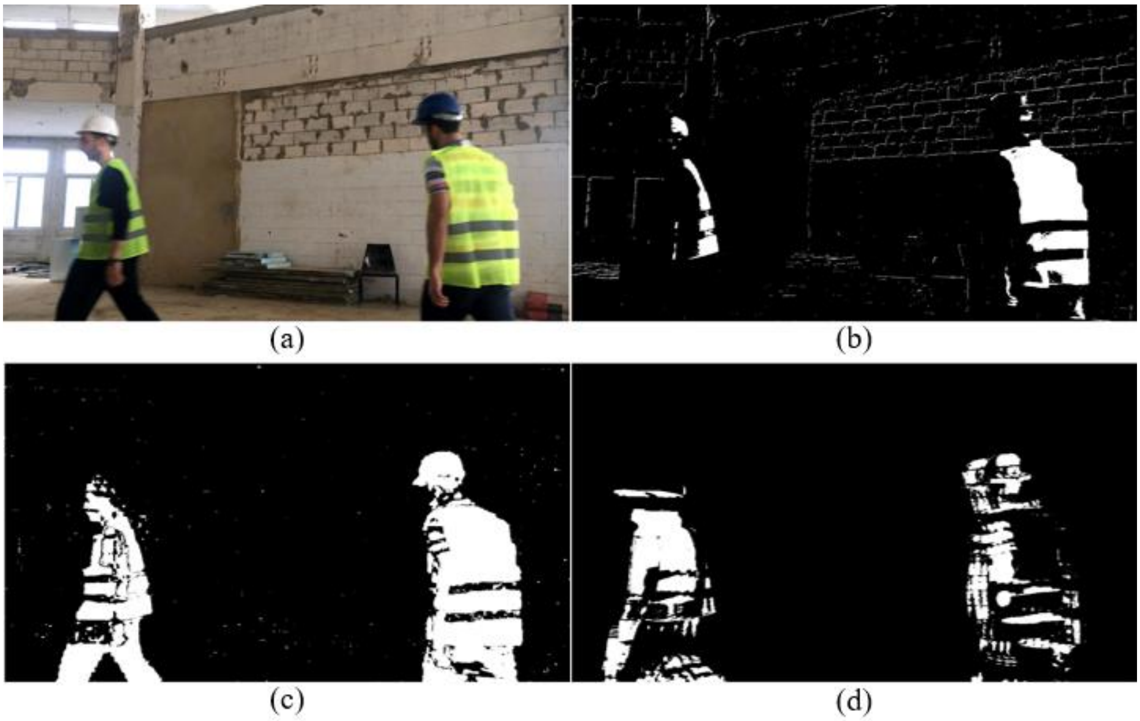

Figure 3.21: Sample image 2: (a) Original image, (b) Background subtraction in RGB, (c) Background subtraction in LAB, and (d) Standard Deviation Method

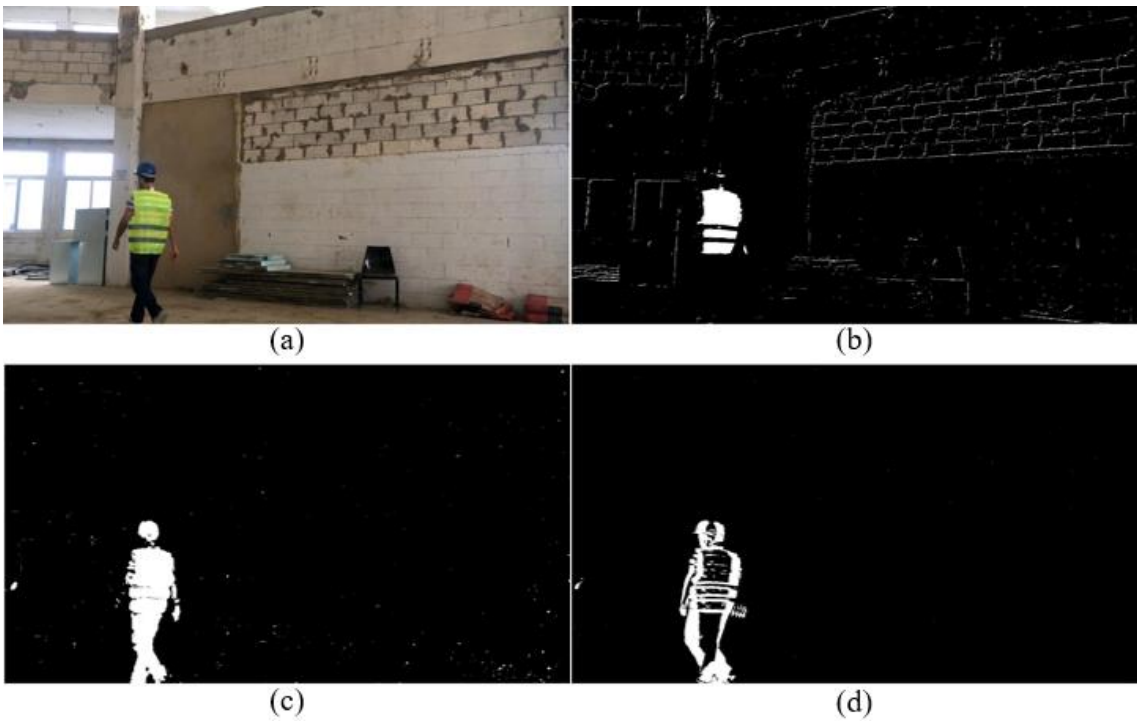

Figure 3.22: Sample image 3: (a) Original image, (b) Background subtraction in RGB, (c) Background subtraction in LAB, and (d) Standard Deviation Method

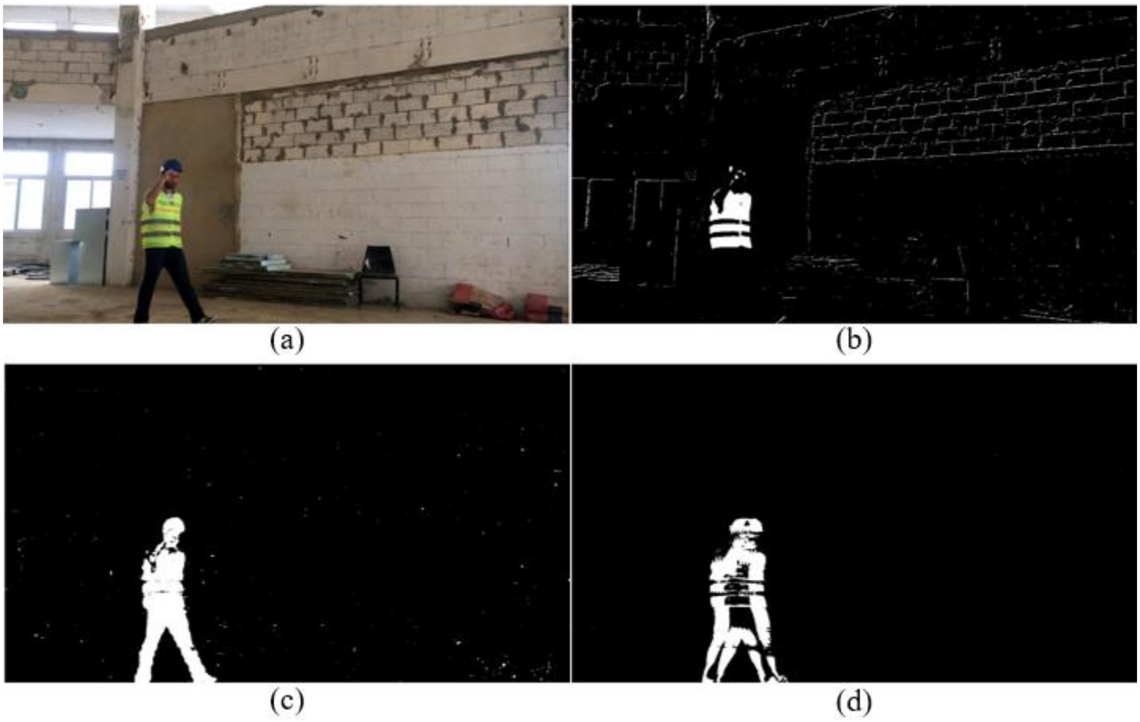

Figure 3.23: Sample image 4: (a) Original image, (b) Background subtraction in RGB, (c) Background subtraction in LAB, and (d) Standard Deviation Method

Using the LAB color space and ignoring differences in luminance offers a noise reduction improvement compared to the RGB color space. On the other hand, the standard deviation method eliminates most of the noise and is robust against small unrepeated color variations. Another advantage of the new proposed method is that it requires no modeling of the background, while the other methods would yield inaccurate results when the background model is not conveniently and actively updated.

In the second set of experiments, the ability of the developed method and the background subtraction using Gaussian Mixture Models in detecting moving components was tested using a dataset of 5 videos. In both methods, similar morphological opperations and blob analysis functions were applied on the obtained binary images to detect movement. Generally, both algorithms were capable of accurately identifying all forms of movement. However, using the standard deviation method, many wrong detections can be potentially avoided. When an object remains in a certain position for a relatively long period, it could be considered within the background model and when it eventually moves, its previous position is mistakenly considered as foreground (Figure 3.24). Since the standard deviation method does not rely on a background model, this case of wrong detection did not occur (Figure 3.25). Wrong detections due to noise are also greatly reduced by using the standard deviation method. In approximately 5 minutes of still video data, the number of false detections did not exceed one digit.

(a) (b)

Figure 3.24: Background Subtraction: (a) Initial detection, and (b) Wrong detection in initially detected region

(c) (d)

Figure 3.25: Standard Deviation Method: (a) Initial detection, and (b) Absence of wrong detection in initially detected region

Another notable advantage is the application of the proposed algorithm on videos captured using a semi-stable camera. This is considered a potential opportunity for adopting infrastructureless, mobile and/or wearable cameras (e.g. handheld phone, drones, etc.) as data acquisition systems. While the accuracy does drop, the system can be optimized to compromise the rate of positive detections in favor of reducing false detections. Detecting construction workers or personnel once every couple of seconds is still sufficient to identify most cases of non-compliance with PPE regulations (Park et al. 2015). Figures 3.26 and 3.27 depict the results generated from both algorithms (i.e. background subtraction and standard deviation) when applied on a short video captured using a hand-held mobile phone.

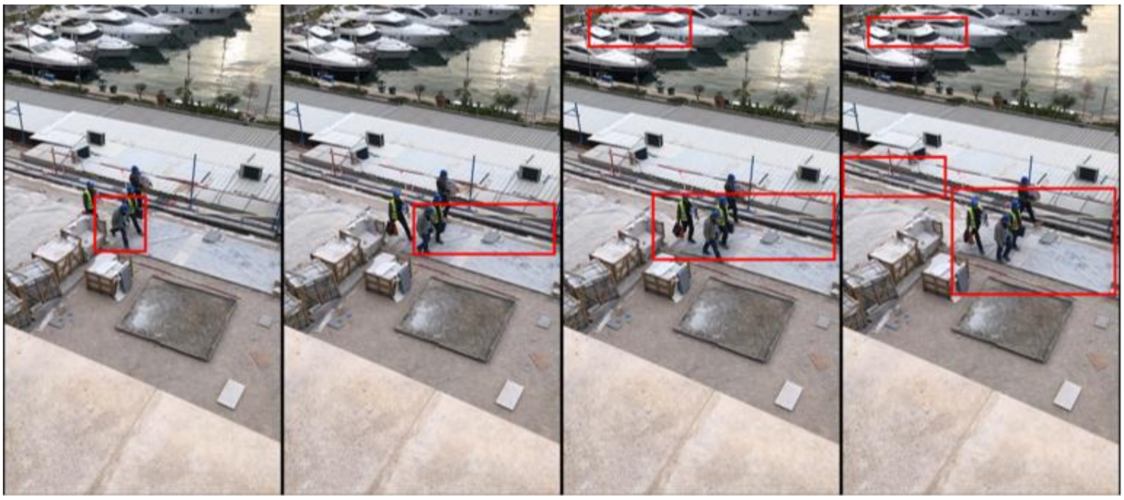

Figure 3.26: Detected motion using GMM background subtraction

Figure 3.27: Detected motion using Standard Deviation method

The proposed method is less susceptible to wrong detections. However, the motion of water was also detected (Figure 3.27). As previously stated, for an outdoor application, the use of a human classifier is required to filter detected movements that do not correspond to on-site individuals.

3.4.2. Evaluation of the hardhat detection algorithm

As aforementioned, the hardhat detection algorithm comprises two main steps: (1) Using a cascade object detector on the upper region of the identified personnel, and (2) Applying color analysis and segmentation to discard wrong detection. A hardhat worn by a worker needs to be inside the defined region of interest to obtain true positive results. On the contrary, a hardhat not worn by a worker needs to be outside the region of interest to avoid false detections. Using two samples of 100 positive and negative instances each, the two previous conditions were always met for all possible worker postures and orientations (Figures 3.28 and 3.29).

Figure 3.28: Hardhats worn by workers – Inside the region of interest

Figure 3.29: Hardhats not worn by workers – Outside the region of interest

Based on these findings, only hardhats actually worn by workers can be identified using the cascade object detector. When the detector mistakenly identifies an irrelevant object as a hardhat, the color segmentation algorithm can still rectify the mistake. However, if the classifier misses an existing hardhat, color segmentation can not detect it.

In order to quantitatively assess the accuracy and precision of the cascade classifier, precision and recall are used as evaluation metrics. The former indicates the percentage of detections that are true positives, while the latter indicates the percentage of true positives that are detected, according to the following formulae:

$$\text{Precision} = \text{True positive} / \text{Total Detections} \tag{3.3}$$
$$\text{Recall} = \text{True Positive} / \text{Total Positive Instances} \tag{3.4}$$

However, in this application, recall is more important than precision and over prediction of hardhats can be tolerated. For a precision value as low as 20%, the system can correctly identify the absence of a hardhat once every five iterations, meaning that a worker not wearing a hard had would be detected in a couple of seconds.

For the aforementioned reasons, a weak classifier was used, with a number of stages equal to 5 and an acceptable false alarm rate of 0.15. Since the value of recall is more important than precision, the cascade object detector was tested using 100 positive sample images containing one hardhat each, with different colors, orientations, background contrast and illumination (Figure 3.30).

Figure 3.30: Sample of testing images

Results indicated that the weak classifier had a low precision of 57.5%. The calculated precision would be lower if negative images were included in the sample, since the detector would be more prone to incorrectly detect portions of the head as hardhats (Mneymneh et al. 2017). The classifier could not detect any hardhat in images 18, 29, 40, 52, and 93, and predicted erroneously the location of the hardhat in images 20, 65, 70, and 78 (table 3.1). In total, 9 hardhats out of 100 were missed, achieving thereby a recall of 91%.

Table 3.1: Cascade object classifier detection results

ID	D*	ID	D	ID	D	ID	D	ID	D	ID	D	ID	D	ID	D	ID	D	ID	D
1	3	11	1	21	2	31	2	41	3	51	1	61	2	71	1	81	1	91	1
2	3	12	1	22	1	32	1	42	1	52	0	62	3	72	1	82	1	92	4
3	2	13	2	23	1	33	3	43	1	53	1	63	3	73	1	83	3	93	0
4	2	14	1	24	4	34	1	44	1	54	2	64	1	74	3	84	1	94	2
5	2	15	3	25	1	35	1	45	2	55	2	65	2**	75	1	85	1	95	2
6	1	16	1	26	2	36	1	46	2	56	1	66	1	76	1	86	1	96	1
7	3	17	2	27	2	37	1	47	1	57	3	67	1	77	1	87	1	97	1
8	2	18	0	28	1	38	1	48	1	58	3	68	2	78	1**	88	2	98	1
9	1	19	1	29	0	39	1	49	1	59	2	69	3	79	4	89	2	99	1
10	2	20	2**	30	2	40	0	50	1	60	1	70	2**	80	3	90	1	100	1

* Number of detected hardhats
** Detections do not include any correct identification of a hardhat

The second step of the algorithm involves applying color segmentation to reduce the number of false identifications. Color segmentation is based on the rules defined in section 4.2, and is applied on regions classified as hardhats. Detections identified in table 3.1 were either confirmed and classified according to their color, or filtered (Figure 3.31). Results from the color segmentation algorithm are very promising whereby most of the wrong detections were filtered and all correct detections were confirmed (Table 3.2). The number of true positives did not change, but the precision was greatly improved from 57.5% to 92%.

Table 3.2: Color segmentation detection results

ID	C*	ID	C*	ID	C*	ID	C*	ID	C*	ID	C*	ID	C*	ID	C*	ID	C*	ID	C*
1	2	11	1	21	1	31	1	41	1	51	1	61	1	71	1	81	1	91	1
2	2	12	1	22	1	32	1	42	1	52	0	62	1	72	1	82	1	92	1
3	1	13	1	23	1	33	2	43	1	53	1	63	1	73	1	83	1	93	0
4	1	14	1	24	1	34	1	44	1	54	2	64	1	74	1	84	1	94	1
5	1	15	1	25	1	35	1	45	1	55	2	65	0	75	1	85	1	95	1
6	1	16	1	26	1	36	1	46	1	56	1	66	1	76	1	86	1	96	1
7	2	17	1	27	2	37	1	47	1	57	2	67	1	77	1	87	1	97	1
8	1	18	0	28	1	38	1	48	1	58	1	68	1	78	0	88	1	98	1
9	1	19	1	29	0	39	1	49	1	59	1	69	1	79	1	89	1	99	1
10	1	20	0	30	1	40	0	50	1	60	1	70	0	80	1	90	1	100	1

* Number of confirmed hardhats

Figure 3.31: Hardhat detection results for positive images, (a) Cascade object detector, and (b) Cascade object detector and color segmentation

To further assess the capability of the method in detecting a failure of wearing PPE, a sample of 30 negative images, containing an individual not wearing a hardhat, were tested (Figure 3.32). Results are presented in table 3.3.

Table 3.3: Color segmentation detection results – negative samples

ID	D*	C**	F**	ID	D	C	F	ID	D	C	F
1	1	0	1	11	1	0	1	21	1	0	1
2	1	0	1	12	2	1	1	22	3	0	3
3	2	0	2	13	2	0	2	23	2	0	2
4	1	0	1	14	2	0	2	24	0	0	0
5	2	0	2	15	1	0	1	25	1	0	1
6	2	0	2	16	6	0	6	26	2	0	2
7	2	1	1	17	1	0	1	27	3	0	3
8	1	0	1	18	0	0	0	28	1	0	1
9	2	0	2	19	0	0	0	29	2	0	2
10	2	0	2	20	1	0	1	30	1	0	1

* Detections using cascade object detector
** Confirmed using color segmentation (False Positives)
*** Filtered using color segmentation (True Negative)

Figure 3.32: Hardhat detection results for negative images, (a) Cascade object detector, and (b) Cascade object detector and color segmentation

The full proposed method was capable of identifying a failure of wearing a hardhat in 28 out of 30 negative images. The weak classifier alone provided only three true negative results. The precision and recall values of the two proposed algorithms can be summarized in table 3.4.

Table 3.4: Precision and recall value for hardhat detector

Method	**Cascade Object Detector**	**Cascade Object Detector and Color Segmentation**
Precision	57.5%	92%
Recall	91%	91%

3.5. Complete framework integration and assessment

The standard deviation moving detection algorithm, the cascade object detector and the color segmentation analysis were integrated into one complete, integrated framework that acquires a video from a construction site, detects and locates motion, searches for hardhats in the upper third of the identified individual using a weak cascade object detector, and filters the obtained results using color segmentation. A warning message is then issued in case a hardhat can not be detected in an area identified as a person. In order to make the system suitable for real-time applications, the detection tool is programmed to be initialized once every 20 frames. As a result, the time interval between two detections is less than a second and the duration required to scan through a video is less than the actual duration of a video. Table 3.5 highlights the time performance of the framework when applied to a video of 55 seconds.

Table 3.5: Time performance of the complete framework

Action	Number of Calls	Total Time	Percentage of Total Time	Average Time
Reading Frame data	82	21.800 s	50 %	0.266 s
Calculating Standard deviations	82	5.476 s	12.5 %	0.067 s
Reshape operations	82	3.326 s	7.6 %	0.040 s
All other operations	82	13.048 s	29.9 %	0.159 s
Total	82	43.650 s	100 %	0.532 s

The system was tested using 4 videos having a total duration of 4 mins and 25 seconds. The overall number of frames is 7960, providing a total of 398 (7960/20) test results. Possible outcomes of the system can be divided into four categories: (1) True warning, (2) False warning, (3) True no warning, and (4) False no warning. Results from the performed experiment are summarized in table 3.6:

Table 3.6: Detection results of complete framework

	Hardhats worn	Hardhats not worn
Warning issued	33 [(a)]	138 [(b)]
Warning not issued	202 [(c)]	25 [(d)]

[(a)] False warning
[(b)] True warning
[(c)] True no warning
[(d)] False no warning

According to the aforementioned experiment, 85% of the total classifications were correct. On the other hand, 80.7% of the total warnings issued were correct (precision), while individuals not wearing hardhats were correctly identified 84.6% of the time (recall). Failing to detect an individual not wearing a hardhat at a rate of 15.4% is tolerable, since the algorithm is likely to rectify any mistake in the next iterations. However, once issued, a false warning cannot be corrected. The calculated values of precision and recall for the integrated framework are lower than the values calculated for the hardhat detection algorithm alone, since some errors are induced by the motion detection algorithm. In our experiment, 23 of the total false warnings were provoked by individuals entering and exiting the scene, making the individual's hardhat outside the algorithm's detection area. The other false warnings were caused by the algorithm's failure to detect an existing hardhat. Generally, incorrect results are dispersed and rarely occur in consecutive frames and iterations. An individual entering or exiting the scene can affect the detection results only once and the algorithm is unlikely to miss an existing individual in two consecutive frames. As a result, incorrect results are predominantly rectified in the following iteration, while correct results are reoccurring (Figures 3.33 and 3.34). For this reason, only issuing a warning when detecting a failure to wear a hardhat in multiple consecutive frames would greatly reduce the rate of false warnings.

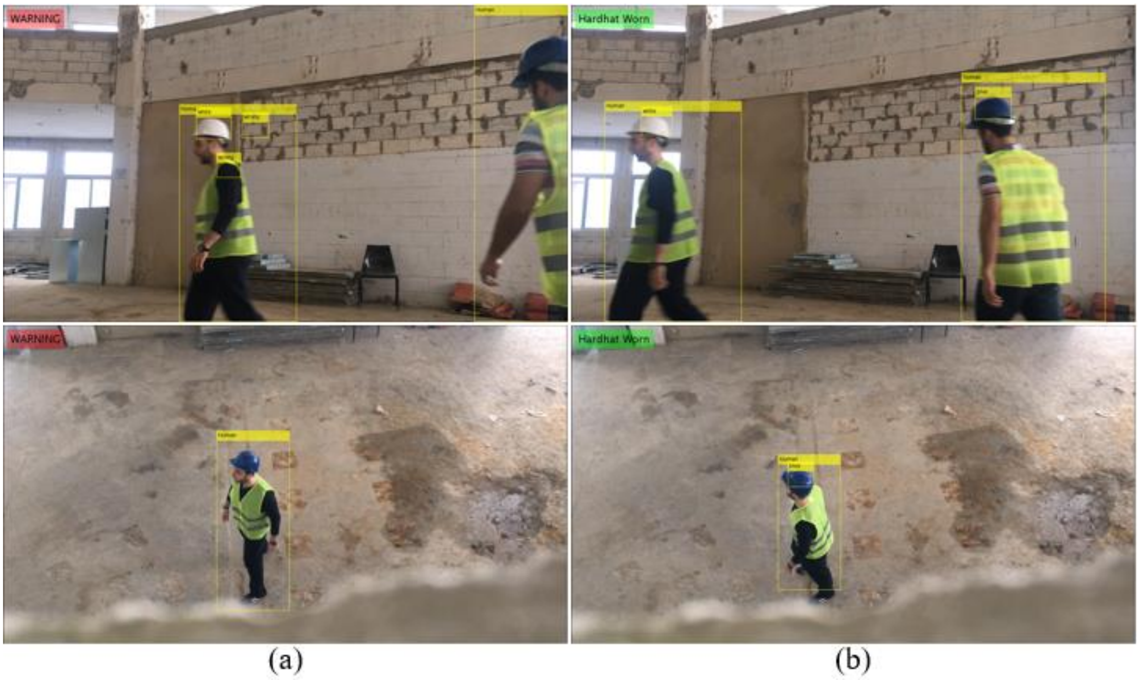

Figure 3.33: (a) Wrong decision in initial frame, and (b) Correct decision in following frame

Figure 3.34: Correct decision in two sets of two consecutive images

Table 3.7: Precision and recall values of the integrated framework

Method	**Integrated Framework**
Precision	80.7%
Recall	84.6%

3.6. Conclusion and future work

This study focused on creating an automated system that can detect compliance with hardhat regulations using computer vision techniques. A warning can then be issued when a region identified as a human does not contain a hardhat. The main contributions of the present study lie in: (1) creating a novel motion detection algorithm based on differences in standard deviations between consecutive images, (2) combining the use of an object detector and color segmentation and analysis for hardhats detection, and (3) integrating the various methods and algorithms into one complete framework. The proposed motion detection algorithm provided more accurate results compared with standard background subtraction methods. A significant noise reduction was achieved and the the need for background modeling was eliminated. This method also displayed potential applicability on videos captured using semi-stabilized cameras. On the other hand, the combination of a cascade object detector together with color analysis and segmentation provided a precision of 92% and a recall of 91% in detecting hardhats. This was achieved by using a weak detector with low precision and high recall, then filtering the obtained results using color segmentation to improve the accuracy of the detector. Finally, the integration of all the proposed methods yielded promising results. For instance, 80.7% of the total warnings issued were correct, while individuals not wearing hardhats were correctly identified 84.6% of the time. All components of the framework functioned smoothly. The system required less time to process a video than its actual duration, making it practical for a real-time application.

Further improvements can be realized. Issuing a warning only when multiple failures of wearing a hardhat are consecutively detected would greatly reduce the rate of false warnings. Other solutions can involve the use of a human classifier so that only individuals entirely visible within each frame are considered. A weaker cascade object detector may also be used to further reduce the number of undetected hardhats, which may also results in an increase in the rate of failure to issue a correct warning. In conclusion, future work is required to improve the accuracy of the system, most importantly in minimizing the rate of incorrect warnings which cannot be rectified, as previously mentioned. A complete framework should ultimately be able to detect various personal protective equipmentand hazardous situations. Additionally, localizing individuals not complying with safety regulations and alarming them accordingly need to be investigated. Finally, future efforts will be channeled towards adopting infrastructureless data acquisition methods.. Autonomous drones could provide a method for surveying large construction areas in a rapid and effective manner.

References

Aranda G. (2000). A study of hazard perception among construction workers: Addressing methodological issues of using navigable movies and repertory grids. *Doctoral Workshop of the Association of Researchers in Construction Management*, Univ. of Salford, ARCOM, 11–15.

Ashworth A. (2013). Contractual Procedures in the Construction Industry. Sixth Edition.

Awwad R., El Souki O., and Jabbour M. (2015). Construction safety practices and challenges in a Middle Eastern developing country. *Safety Science*, 83, 1–11.

Brilakis I., Park M.W., and Jog G. (2011). Automated vision tracking of project related entities. *Advanced Engineering Informatics*. 25 (4) 713-724.

CBC News (2012). 3 most dangerous job sectors in Canada. Available online: http://www.cbc.ca/news/canada/story/2012/04/25/f-dangerous-jobs.html

Chen Z., Li H., and Wong C. T. C. (2000). Environmental management of urban construction projects in China. *Journal of Construction Engineering and Management*, 126 (4), 320–324.

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).

Cheng T., Venugopal M., Teizer J., Vela P. A. (2011). Performance evaluation of ultra wideband technology for construction resource location tracking in harsh environments. *Automation in Construction,* 20(8), 1173–1184.

Chi S., Caldas C. H., and 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, 199-211.

Chiocha C., Smallwood J., Emuze F. (2011). Health and safety in the Malawian construction industry. *Acta Structilia*. 18 (1), 68–80.

Coble R. J., Hinze J., and Haupt T. C. (2000). Construction safety and health management, *Prentice-Hall, Englewood Cliffs, N.J*.

Dimitrov A. and 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*, 28 (1), 37 – 49.

Ding L., Yu H. L., Zhou C., Wu X., and Yu M. H. (2010). Safety risk identification system for metro construction on the basis of construction drawings. *Automation in Construction,* 27, 120-137.

Du S., Shehata M. and Badawy W. (2011). Hardhat Detection in Video Sequences based on Face Features, Motion and Color Information. *3rd International Conference on Computer Research and Development*.

Elgammal A., Hanvood D., and Davis L.S. (2000). Nonparametric model for background subtraction. European Conference on Computer Vision, 751-767.

Gheisari M., Irizarry J., and Walker B. N. (2014). UAS4SAFETY: The Potential of Unmanned Aerial Systems for Construction Safety Applications. *Construction Research Congress*.

Hamledari H., McCabe B., and Davari S. (2017). Automated computer vision-based detection of components of under construction indoor partition. *Automation in construction*, 74, 78 – 94.

Han B., Comaniciu D., and Davis L.S. (2004). Sequential kemel density approximation through mode propagation: applications to background modeling. *Asian Conference on Computer Vision*, 2004.

Health and Safety Executive (2014). Health and safety in construction sector in Great Britain, 2014/15. Available online: http://www.hse.gov.uk/statistics/industry/construction/construction.pdf

Hinze J., Huang X., and Terry L. (2005). The nature of struck-by accidents. *Journal of Construction Engineering and Management*, 131 (2), 262–268.

Hinze J., Pedersen C., and Fredley J. (1998). Identifying root causes of construction injuries. *Journal of Construction Engineering and Management*, 124 (1), 67–71.

Hunter R. S. (1948). Photoelectric Color-Difference Meter. *Proceedings of the Winter Meeting of the Optical Society of America*, 38 (7), 661.

Hwang S. (2012). Ultra-wide band technology experiments for real-time prevention of tower crane collisions. *Automation in Construction*, 22, 545–553.

Joblove G. H. and Greenberg D. (1978). Color spaces for computer graphics. *Computer Graphics*, 12 (3): 20–25.

Kasirossafar M. and Shahbodaghlou F. (2013). Application of visualization technologies to design for safety concept. *Gateway to a Better Tomorrow – Proceedings of the 6th Congress on Forensic Engineering*, San Francisco.

Kim Y. S., Lee J. H., Yoo H. S., Lee J. B. and Jung U. S. (2009). A performance evaluation of a Stewart platform based Hume concrete pipe manipulator. *Automation in Construction*, 18 (5), 665-676.

Lee G., Cho J., Ham S., Lee T., Yun S. H., and Yang H. J. (2012). A BIM- and sensor-based tower crane navigation system for blind lifts. *Automation in Construction*, 26, 1 – 10.

Lienhart R. and Maydt J. (2002). An extended set of Haar-like features for rapid object detection. *Proc. of IEEE International Conference of Image Processing*, 1, 900-903.

Lingard H., and Rowlinson S. (1997). Behaviour-based safety management in Hong Kong's construction industry. *Journal Safety Research*, 28 (4), 243–256.

Liu C., Sharan L., Adelson E.H., and Rosenholtz R. (2010). Exploring features in a bayesian framework for material recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 239–246.

Lo B.P.L., and Velastin S.A. (2001). Automatic congestion detection system for underground platforms. *Proceedings of International Symposium on Intelligent Multimedia, Video and Speech Processing*, 158-161.

Marks E., Teizer J. (2013). Evaluation of the position and orientation of (Semi-) passive RFID tags for the potential application in ground worker proximity detection and alert devices in safer construction equipment operation. *ASCE International Workshop on Computing in Civil Engineering*, 645–652.

MathWorks. Vision.ForegroundDetector System object. available at: https://www.mathworks.com/help/vision/ref/vision.foregrounddetector-class.html?searchHighlight=foreground%20detector&s_tid=doc_srchtitle

Memarzadeh M., Golparvar-Fard M., and 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,

Occupational Safety and Health Administration (2006). Module 13 - Personal protective equipment.

Oliver N.M., Rosario B., and Pentlan A.P. (2000). A Bayesian computer vision system for modeling human interactions. *IEEE Transactions on Paftern Analysis and Machine Intelligence*, 22 (8), 831-843.

Park M.W., Elsafty N., and Zhu Z. (2015). Hardhat-Wearing Detection for Enhancing On-Site Safety of Construction Workers. *Journal of Construction Engineering and Management*.

Park M.W., Koch C., and 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.

Park M.W., Palinginis E., and Brilakis I. (2012). Detection of construction workers in video frames for automatic initialization of vision trackers. *Construction Research Congress*. 940 – 949.

Piccardi M. (2004). Background subtraction techniques: a review. *IEEE International Conference on Systems, Man and Cybernetics*, 2004.

Rubaiyat A.H.M., 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. *IEEE/WIC/ACM International Conference on Web Intelligence Workshops,* 135 – 142.

Seki M., Wada T., Fujiwara H., and Sumi K. (2003). Background subtraction based on cooccurrence of image variations. *IEEE Conference on Computer Vision and Pattern Recognition*, 2, 65-72.

Seo J., Han S., Lee S. and Kim H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, 29, 239-251.

Sherratt F., Farrell P., and Noble R. (2013). UK construction site safety: discourses of enforcement and engagement. *Construction Management and Economics*, 31 (6), 623-635.

Shrestha K., Sherestha P. P., Bajracharya D., and Yfantis E. A. (2015). Hard-Hat Detection for Construction Safety Visualization. *Journal of Construction Engineering*.

Skibniewski, M. J. (2014). Information technology applications in construction safety assurance. *Journal of civil engineering and management*, 20 (6), 788-794.

Stauffer C., and Grimson W.E.L. (1999). Adaptive background mixture models for real-time tracking. *IEEE Conference on Computer Vision and Pattern Recognition*, 246-252.

Teizer J., Castro-Lacouture D. (2007). Combined ultrawideband positioning and range imaging sensing for productivity and safety monitoring in building construction. *Congress on Computing in Civil Engineering*, 681–688.

Teo A. L., Theo H., Feng Y. (2008). Construction health and safety performance in developing and developed countries: a parallel study in South Africa and Singapore. *In: Hinze, Jimmie (Ed.), Evolution of and Directions in Construction Safety and Health.*

United States Department of Labor, Bureau of Labor Statistics (2015). National Census of Fatal Occupational Injuries in 2015. Available online: https://www.bls.gov/news.release/pdf/cfoi.pdf.

United States Department of Labor, Occupational Safety and Health Administration (2015). Commonly Used Statistic. Available online: https://www.osha.gov/oshstats/commonstats.html.
Viola P. and Jones M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*. 57 (2), 137-154.
Wren C., Azarhayejani A., Darrell T., and Pentland, A.P. (1997). *Pfinder: real-time tracking of the human body*. IEEE Transactions on Pattern Analysis and Machine Intellignece, 19 (7), 780-785.
Yan J., Park, M.W., Vela P.A., and Golparvar-Fard M. (2015). Construction performance monitoring via still images, time-lapse photos, and video streams: now, tomorrow, and the future. *Advanced Engineering Informatics*, 29 (2) 211-224.
Yoshida T. and Chae S. (2010). Application of RFID technology to prevention of collision accident with heavy equipment. *Automation in Construction*, 19 (3), 368-374.