2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops

Automatic Detection of Helmet Uses for Construction Safety

Abu H. M. Rubaiyat[1], Tanjin T. Toma[1], Masoumeh Kalantari-Khandani[1],
Syed A. Rahman[1], Lingwei Chen[1], Yanfang Ye[1*] , Christopher S Pan[2]
[1]Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV 26506
[2]Division of Safety Research, National Institute for Occupational Safety and Health, Morgantown, WV 26505
[1]{abrubaiyat, tatoma, makalantarikhandani, srahman2, lgchen}@mix.wvu.edu, yanfang.ye@mail.wvu.edu, [2]syp4@cdc.gov

***Abstract*—The U.S. construction industry suffers from the highest number of fatalities among all industries, i.e., one in five worker deaths in private industry were in construction. Tremendous loss has occurred to the workers' families, the industry, and the nation. Considering the large and increasing number of construction projects that are being conducted in the U.S., there is a growing necessity of developing innovative methods to automatically monitor the safety for the workers at construction sites. Since the head is the most critical area of a human body and is the most vulnerable to an impact that could cause serious injury or death, the use of a protective helmet in construction work is needed. In this paper, we aim to automatically detect the uses of construction helmets (e.g., whether the construction worker wears the helmet or not) by analyzing the construction surveillance images. Based on the collected images, we first detect the object of interest (i.e., construction worker) and further analyze whether the worker wears the helmet or not, by using computer vision and machine learning techniques. In the first step, we incorporate frequency domain information of the image with a popular human detection algorithm Histogram of Oriented Gradient (HOG) for construction worker detection; in the second step, the combination of color-based and Circle Hough Transform (CHT) feature extraction techniques is applied to detect helmet uses for the construction worker.**

I. Introduction

In the United States, many people work at jobsites under unsafe conditions, and thousands lose their lives every year. Actually, the U.S. construction industry suffers from the highest number of fatalities among all industries, i.e., one in five worker deaths in private industry in 2014 were in construction [35]. To put this into perspectives, the number of worker deaths in construction (9,836 in 2005-2014) is even 44% more than the American war and military operations fatalities (6,830 in 2001-2014) in the past decade [8]. Tremendous loss has occurred to the workers' families, the industry, and the nation: the average of fatal occupational injuries in construction would represent a loss of $5.2 million [27]. To protect the nation's construction workforce, methods to improve safety performance measurement on construction sites is of paramount importance [17].

The causes of the construction site fatalities include falls, slips, being struck by objects, electrocution, and being caught in/between objects [25]. And falls to a lower level are the leading hazards that have caused construction fatalities, accounting for one third of work deaths on construction sites [6]. In most of the fall incidents, the workers fall from heights and hit their heads on hard floors. In one study that investigated the number of construction fatalities and the use of safety equipment, the results showed that 47.3% of fatally injured victims either had not used safety equipment (e.g., helmet, guard rails, etc.) or had not used them properly [1]. Since the head is the most critical area of a human body and is the most vulnerable to an impact that could cause serious injury or death, the use of a protective helmet in construction work is required. However, the construction workers would not always follow the Occupational Safety and Health Administration (OSHA) regulations to wear head protection (e.g., helmet) whenever OSHA regulations require that they do so (e.g., under conditions of elevation). Therefore, methods to improve safety performance measurement on construction sites is of paramount importance [17]. Considering the large and increasing number of construction projects that are being conducted in the U.S. [7], there is a growing necessity of developing innovative methods to automatically monitor the safety for the workers at construction sites. Thanks to the widespread use of mobile sensors and new emerging sensor technologies, as well as the availability of data on various aspects of job bidding, construction equipment usage, and other data-driven applications, visual data surveillance on construction sites is exploding, and we have entered the era of big data construction. Surveillance of construction safety is now becoming more data driven [8]. In this paper, we aim to automatically detect the uses of construction helmets (e.g., whether the construction worker wears the helmet or not) by analyzing the construction surveillance images. Based on the collected images, we first detect the object of interest (i.e., construction worker) and further analyze whether the worker wears the helmet or not, by using computer vision and machine learning techniques.

Detection of construction worker with or without safety equipment (i.e., helmet) in construction surveillance images leads to the identification of safety violations. Figure 1 shows two cases, where Figure 1 (a) illustrates the positive example (construction worker with helmet) and Figure 1 (b) indicates the negative example (construction worker without helmet). In this paper, to automatically detect helmet uses for construction

* Corresponding author

978-1-5090-4771-0/16 $31.00 © 2016 IEEE
DOI 10.1109/WIW.2016.10

safety, based on the collected construction surveillance images, we first use Discrete Cosine Transform (DCT) and Histogram of Oriented Gradient (HOG) features which are fed to Support Vector Machine (SVM) for object of interest (i.e., construction worker) detection. Afterwards, the combination of color-based and Hough Transform feature extraction techniques is applied to detect helmet uses for the construction worker.

(a)

(b)

Fig. 1: Examples of helmet uses in construction sites (Data source: Google Images)

This work is innovative, in that it combines the emerging computer vision and machine learning techniques to create a collaborative platform for construction safety performance measurement that helps to reduce construction worker fatalities and serious injuries caused by falls to a lower level. The prototype developed in the paper is a first-of-its-kind system that allows the stakeholders (e.g., contractors, architects, engineers, builders and owner representatives) to monitor and detect the uses of helmets on construction sites.

The rest of this paper is organized as follows. Section II discusses the related work. Section III introduces the developed system architecture and Section IV describes the proposed method in detail. Section V systematically evaluate the performance of our proposed method. Finally, Section VI concludes.

II. Related Work

A. Construction Worker Detection

The first step of our work is construction worker detection from the collected construction surveillance images. The problem of human (e.g., construction worker) detection is to automatically locate people in an image or video sequence, which has been actively investigated in the past decade. Human detection has variety of applications such as video-based surveillance, automatic tagging in visual content management, autonomous driving [23], etc. The problem of human detection has many challenges associated with it. The non-rigid nature of the human body produces numerous possible poses. It is also challenging to model simultaneously view (orientation) and size variations arisen from the change of the position and direction (e.g. tilt angle) of the camera. Unlike other types of objects, humans can be clothed with varying colors and texture, which adds another dimension of complexity. Furthermore, a significant percentage of scenes, such as urban environments, contain substantial amounts of clutter and occlusion [30].

Currently, the most prevalent approaches presented in the literatures are the detector-style methods, in which detectors are trained to search for humans within an image or video sequence over a range of scales. A number of these methods use global features such as Histogram of Oriented Gradient (HOG) descriptor [7], edge templates [12], while others build classifiers based on local features such as SIFT-like descriptors [22], Haar wavelets [36], and SURF-like descriptors [16]. Another family of approaches models humans as a collection of parts [21], [28], [31]. Typically this class of approaches relies on a set of low-level features which produce a series of part location hypotheses. Subsequently, inferences are made with respect to the best assembly of existing part hypotheses. Approaches such as AdaBoost have been used with some degree of success to learn body part detectors such as the face [37], hands, arms, legs, and torso [21], [29]. A considerable amount of works have also focused on shape based detection. Zhao et al. [41] used a neural network that was trained on human silhouettes to verify whether the extracted silhouettes correspond to a human subject. However, a potential disadvantage of this method resides in the fact that they relied on depth data to extract the silhouettes. Others, such as Davis et al. [42] have also attempted to make use of shape-based cues by comparing edges to a series of learned models. Wu et al. [39] have proposed learning human shape models and representing them via a Boltzmann distribution in a Markov Field.

Although a number of these methods have proved to be successful in detecting humans in the images, we have considered HOG descriptors because of their simple structure and high performance in human (e.g., construction worker) detection.

B. Helmet Use Detection

The literature of helmet use detection is very limited. It is considerably a new topic in computer vision and machine learning. Majority of the works focused on using color information for helmet detection. Du et al. [10] described a combined machine learning and image processing approach for helmet detection in video sequences. In their framework, there were three major parts: the first was the person's face detection based on Haar-like face features [20]; the second was the motion detection and skin color detection used to reduce the false alarms of faces; the third was the helmet detection using the color information above the face regions. For both the face detection and the helmet detection, they used the YCbCr [19] and HSV [32] color spaces. In a similar work, Park et al. [26] exploited HOG features for human body detection and subsequently used color histograms for helmet detection. In another work, Wen et al. [38] proposed a circle detection method called Modified Hough Transform for helmet detection for ATM's surveillance systems.

In this work, we will explore to combine color-based and Circle Hough Transform (CHT) feature extraction techniques in order to develop a more robust and accurate helmet use detection system.

III. System Architecture

The overall system architecture for helmet use detection for construction safety is performed based on the construction surveillance images, which consists of three major components: image segmentation, object of interst (i.e., construction worker) detector, and helmet use detector, as illustrated in Figure 2.

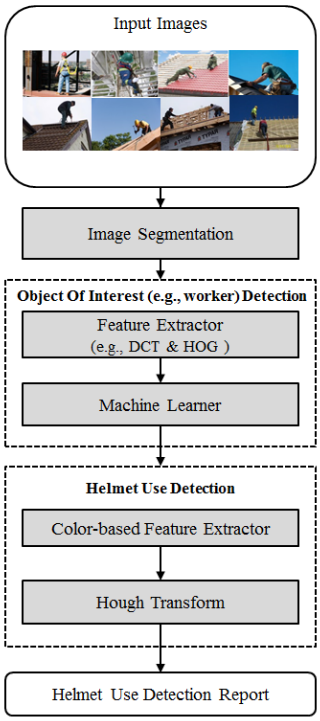


Fig. 2: System architecture of helmet use detection

- *Image Segmentation:* For the collected images, a semantic image segmentation algorithm, such as Gaussian Mixture Model (GMM), is first applied to partition each of the relevant construction surveillance images into a set of object regions (e.g., scaffold, roof, sky, worker, etc).
- *Object of Interst Detector:* After image segmentation, in order to recognize whether the segmented object regions are construction workers, Discrete Cosine Transform (DCT) is computed to extract the frequency domain information from the spatial domain image, and then Histogram of Oriented Gradient (HOG) features are drawn from the DCT coefficients. Resting on these features of the segmented regions, supervised classifier (i.e., Support Vector Machine (SVM) with linear kernel) is applied to detect whether there's construction worker in the image. (See Section IV-B for detail.)
- *Helmet Use Detector:* After detecting the object of interest (i.e., construction worker in our application), a combination of color-based and Circle Hough Transform (CHT) feature extraction techniques is applied for helmet use detection. (See Section IV-C for detail.)

IV. Proposed Method

A. Problem Definition

Based on the collected construction surveillance images, after image segmentation (in our application, we use Gaussian Mixture Model (GMM) for image segmentation), we represent our dataset $D = \{\mathbf{x}_i, y_i, z_i\}_{i=1}^n$ of n segmented images, where $\mathbf{x}_i$ is the set of features extracted from the segmented image i, y_i is the class label of image i where $y_i \in \{human, non\text{-}human\}$, and z_i is the class label of image i where $z_i \in \{with-helmet, without-helmet\}$. Let d be the number of features, and then $\mathbf{x}_i \in \mathbb{R}^d$.

The helmet detection problem can be specified as follows: given a dataset D as defined above, assign a label y (i.e., *human* or *non-human*) to an input image $\mathbf{x}$ through a classifier f; for the images with *human* labels, further assign a label z (i.e., $with-helmet$ or $without-helmet$) to each of them. Accordingly, in this paper, the proposed method can be divided into two steps: (1) construction worker detection, and (2) helmet use detection. In the first step, Discrete Cosine Transform (DCT) is used to extract frequency domain information from the segmented images and then Histogram of Oriented Gradient (HOG) features are extracted from the DCT coefficients. To predict whether construction worker is included in the image, the state-of-the-art supervised classifier Support Vector Machine (SVM) with linear kernel is used. After detecting the objects of interest (i.e., construction worker in our application), a combination of color-based and Circle Hough Transform (CHT) feature extraction techniques is exploited. Based on the color and shape information, the proposed method detects whether the construction worker wears helmet or not.

B. Construction Worker Detection

1) Discrete Cosine Transform: The Fourier transform decomposes a signal into its sine (imaginary) and cosine (real) components. The real part of the transform actually forms the Discrete Cosine Transform (DCT). The equation of 2D-DCT given by [24] is,

$$D(u,v)=\begin{cases} \frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y), \ if \ \ u=v=0 \\ \frac{1}{N^2}\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)\cos(\frac{(2x+1)u\pi}{2N})\cos(\frac{(2y+1)v\pi}{2N}), \ otherwise \end{cases} \tag{1}$$

where $f(x, y)$ is a discrete valued image with the size of $N \times N$ and $D(u, v)$ is the corresponding 2D version of DCT. This transform is used to compute the projection of an image into the orthogonal basis of cosine functions, resulting in a set of coefficients that represents the image in the real part of the spectral domain. In an image, a huge portion of signal energy lies in the low frequencies which appear in the upper left corner of corresponding DCT. From DCT of an image, distribution of energies in frequency domain can be found. This distribution should be different for human and non-human segments. Using Histogram of Oriented Gradient (HOG), this difference in distribution is further measured.

2) Histogram of Oriented Gradient: The Histogram of Oriented Gradient (HOG) human detector is one of the most popular and successful "human detectors". It was introduced by Dalal and Triggs in [7]. HOG uses a "global" feature to describe a human rather than a collection of "local" features. This means that the entire human is represented by a single feature vector, as opposed to many feature vectors representing smaller parts of the human. HOG human detector uses a sliding detection window which is moved around the image. At each position of the detector window, a HOG descriptor is computed for the detection window. This descriptor is then shown to a trained classifier, which classifies it as either "human" or "non-human".

In this paper, HOG features are computed for the 128×128 detection window. First, the gradient vector is computed at each pixel (both magnitude and angle) for this image segment. This 128×128 image segment is then divided into 16×16 blocks with 50% overlapping. Further, each block is divided in four 8×8 cells. Then, the gradient vectors in each cell are put in a 9-bin (0-180 degrees) histogram. Note that $L2$ normalization method is used for normalizing the histogram to make it invariant to the illumination change. To further illustrate, the 128×128 pixel detection window is divided into 15 blocks horizontally and 15 blocks vertically, for a total of 225 blocks. Each block contains 4 cells with a 9-bin histogram for each cell, for a total of 36 values per block. This brings the final vector size to 15 blocks horizontally $\times$ 15 blocks vertically $\times$ 4 cells per block $\times$ 9-bins per histogram = $8,100$ values. Figure 3 demonstrates the general HOG implementation scheme step by step.

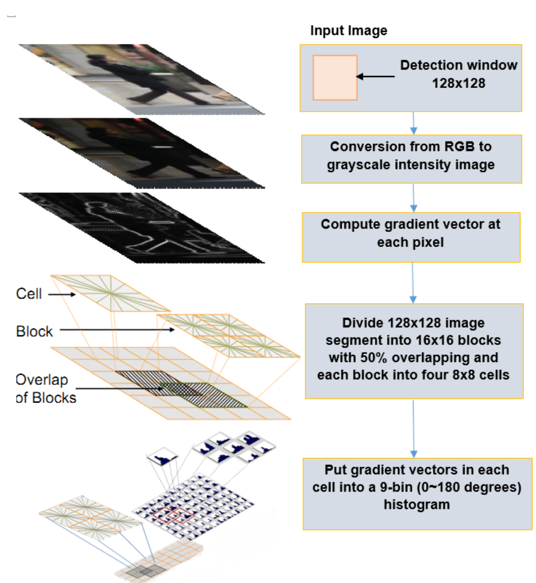


Fig. 3: HOG implementation scheme

3) Support Vector Machine: Support Vector Machine (SVM) is a method for the classification of both linear and nonlinear data [18]. It uses a nonlinear mapping to transform the original training data into a higher dimension. Within this new dimension, it searches for the linear optimal separating hyperplane (i.e., a "decision boundary" separating the data points of one class from another). With an appropriate non-linear mapping to a sufficiently high dimension, data from two classes can always be separated by a hyperplane. The SVM finds this hyperplane using support vectors ("essential" training data points) and margins (defined by the support vectors). SVM can be of linear and non-linear kernels. In our application, we apply linear SVM to classify two classes (human and non-human) due to its high efficiency. The output of a linear SVM is $u = w \times x - b$, where w is the normal weight vector to the hyperplane and x is the input vector. Maximizing the margin can be seen as an optimization problem:

$$\text{minimize} \frac{1}{2}\|w\|^2, \text{ subject to } y_i(w \cdot x + b) \geq 1, \forall i, \quad (2)$$

where x is the training example and y_i is the correct output for the i_{th} training example.

Figure 4 shows the detection flow of the construction worker. After image segmentation, DCT coefficient matrix of an image is used instead of RGB image as the input to HOG features extraction scheme. Then, SVM classifier is trained with the HOG features extracted from human and non-human image blocks. Finally, this trained classifier is used to detect the object of interest (i.e., construction worker) in testing images. The implementation of the proposed construction worker detection method is given in Algorithm 1.

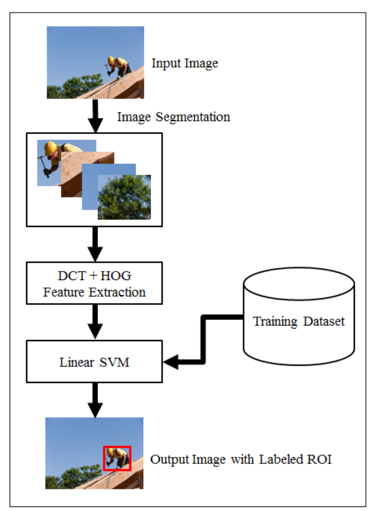


Fig. 4: Construction worker detection flow

C. Helmet Use Detection

1) Color-based Feature Extraction: After object of interest (i.e., construction worker) detection, we aim at searching for helmet use in the image segment to identify safety violation. In most of the construction surveillance images, it can be noticed that certain colors are most frequently used for helmets, such

Input: $D = \{\mathbf{x}_i, y_i\}_{i=1}^{n}$: training image set of n training image samples; $D_t = \{\mathbf{x}_i, y_i\}_{i=1}^{n_t}$: testing image set of n_t testing image samples.
Output: The labels of all testing images: human or non-human.

Train a SVM classifier $f(\mathbf{X})$ using n training image samples;
Partition images into a set of object regions $\mathcal{R}$;
for *each object region* $i \in \mathcal{R}$ **do**
    for *each pixel* (x, y) *in* i **do**
        Using Eq. 1 to calculate 2D-DCT $D(u, v)$;
    end
    Calculate HOG features $\mathbf{x}_i$ using DCT matrix of i;
end
Using the classifier $f(\mathbf{X})$ to detect construction worker in D_t;

Algorithm 1: The algorithm for construction worker detection

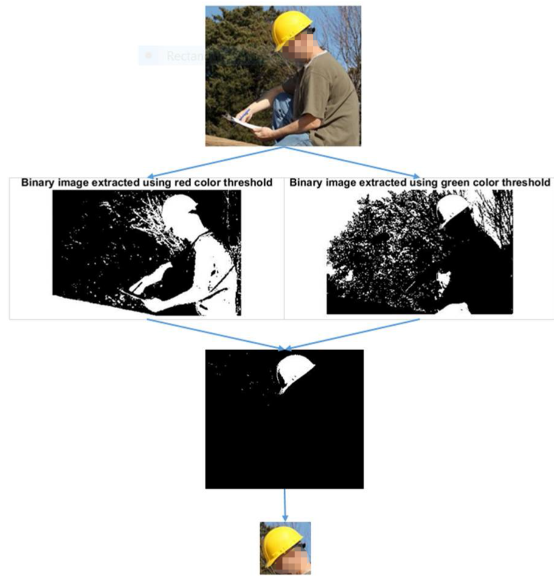


Fig. 5: Detection of helmet with yellow color

as yellow, blue, red and white. Based on this observation, the proposed system is designed to recognize helmets made of these particular colors.

In the color-based feature extraction, threshold based color segment detection is used. For red and blue helmet detection, thresholds for only red and blue colors are set respectively. But for yellow color detection, thresholds for both red and green colors are required. Blue is not dominant as red and green in yellow color. Binary images are generated from red and green color planes using thresholds. Then common region in these two binary images is extracted, which belongs to yellow region. At last stage color information is retrieved for this region from the original RGB image. For white color detection, a common threshold for all three color components (red, green and blue) are used. Figure 5 shows an example of yellow color helmet detection using color-based feature extraction.

Our proposed algorithm searches for one of the four aforementioned color regions in the detected object of interest (i.e., construction worker) sequentially. Once it detects a particular color regions, it computes Hough Transform to find circles in those regions (introduced in the following section). If any circle is detected, it is considered as a helmet.

2) Circle Hough Transform: In general, Hough Transform is a voting scheme to detect certain shapes in images such as lines, squares, circles, etc. In fact, it is a feature extraction technique used in image analysis, computer vision, and digital image processing. The purpose of Circle Hough Transform (CHT) is to find possible circular shapes in images [40]. The circle candidates are produced by "voting" in Hough parameter space. Then the local maxima in a matrix of candidates is picked. If (a, b) is the center and r is the radius of a circle, then the circle can be defined by the following equation:

$$(x-a)^2 + (y-b)^2 = r^2 \tag{3}$$

In CHT, at first the image is converted to binary (black and white) using an edge detection technique such as canny edge detector [9]. The next step is to find some points that are candidate for the centers of the circles for a given radius. Now if there are many radii (smaller than the first) for that fixed point, then there will be several nested circles inside this circle. The system proposed in this paper uses CHT to detect the circle shape around a helmet. After color-based feature extraction, it tries to find circle shape in the image segment. First the diagonal length d of the image segment is calculated using Pythagorean Theorem. Then a percentage of the diagonal length is considered as a range of radii. Maximum ($R_{\max}$) and minimum ($R_{\min}$) of this range are measured using the following equations:

$$R_{\max} = ceil(0.80 \times d) \tag{4}$$

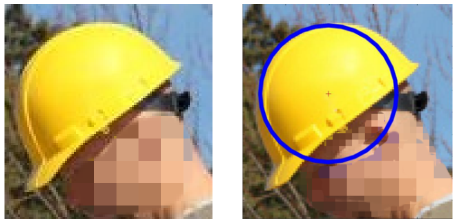
$$R_{\min} = ceil(0.06 \times d) \tag{5}$$

where the values 0.80 and 0.06 for $R_{\max}$ and $R_{\min}$ are found empirically. Then, all the circles that fall within $R_{\min}$ and $R_{\min}$ will be marked. Figure 6 (a) shows the extracted color region from the detected construction worker and (b) shows the detected circle of the helmet in that segment using CHT.

Figure 7 shows the overall flow of helmet use detection. And the implementation of the proposed helmet use detection method is given in Algorithm 2.

V. Experimental Results and Analysis

In this section, to empirically validate the proposed method, we conduct two sets of experiments based on the collected image sample set described in Section V-A: (1) In the first set of experiments, we compare our proposed method for human (i.e., construction worker) detection with the method using



(a) (b)

Fig. 6: Helmet circle detected by CHT

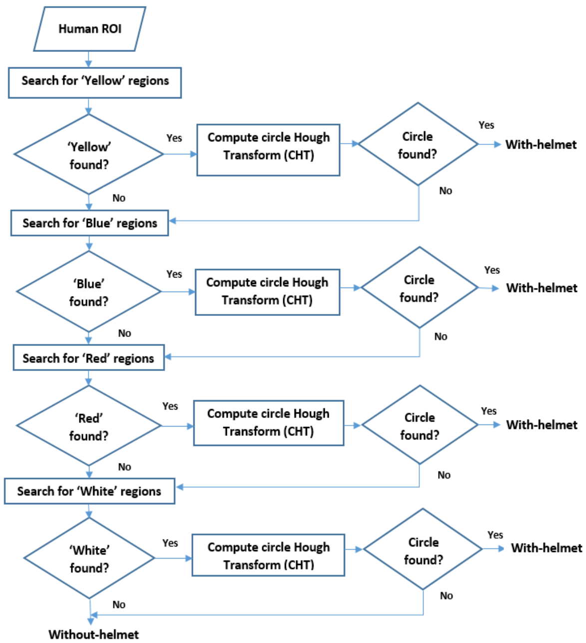


Fig. 7: Helmet use detection flow

Input: $D_t = \{\mathbf{x}_i, y_i, z_i\}_{i=1}^{n_t}$: n_t image segments with detected construction worker (i.e., $y_i = human$ for each image i).
Output: The labels for the testing images: with or without helmet.

for *each image in* D_t **do**
  Calculate d using Pythagorean Theorem;
  Calculate $R_{\min}$ and $R_{\max}$;
  Apply color-based method to extract color region c;
  switch(c);
  case "Yellow" ;
  Compute CHT to find circles with radius r;
  if $r \in (R_{\min}, R_{\max})$ **then**
    return "with-helmet";
  end
  case "Blue" ;
  Compute CHT to find circles with radius r;
  if $r \in (R_{\min}, R_{\max})$ **then**
    return "with-helmet";
  end
  case "Red" ;
  Compute CHT to find circles with radius r;
  if $r \in (R_{\min}, R_{\max})$ **then**
    return "with-helmet";
  end
  case "White" ;
  Compute CHT to find circles with radius r;
  if $r \in (R_{\min}, R_{\max})$ **then**
    return "with-helmet";
  end
  default return "without-helmet";
end

Algorithm 2: The algorithm for helmet use detection

HOG only; (2) In the second set of experiments, we further assess the effectiveness of our proposed helmet use detection method by comparison with the method merely using CHT.

A. Experimental Setup

In data collection stage, the construction images are collected from different websites ([13], [14], [3], [4], [33], [34], [15], [5]). As manual image collection would be time consuming, an image crawler is built to automatically collect images from a given website. We develop a crawler that extracts the source codes from the URL of the website and searches for some key words, basically some image extensions like ".jpg", ".png" etc. Then it extracts the image URL that contains the target key words, and download the corresponding images. The developed crawler downloads all the images found in the given websites including both construction images and some other unnecessary images. At data cleaning stage, the unwanted images are filtered manually. Primarily, around $10,000$ images are collected. After performing data cleaning, $1,000$ images are selected for further experiments.

To train the classifier for human (i.e., construction worker) detection, 354 human and 600 non-human sample images are extracted from the dataset. To prepare this training set, based on the collected construction images, Gaussian Mixture Model (GMM) is exploited for image segmentation. Human samples are of different poses, as in construction site images workers are found to be in different poses based on what they are doing. Non-human samples mainly comprise construction tools, buildings, roofs, sky, and trees etc. that are typically found in construction images. For testing, we further collect 200 construction images, 67 of which are tagged as "with-helmet", 83 are "without-helmet" and 50 are tagged as "non-human". The collected data is described in Table I. We evaluate the performance of different methods using the measures shown in Table II.

B. Evaluation of Construction Worker Detection

For human (i.e., construction worker) detection, HOG features extracted from DCT coefficients of the images are fed to the linear SVM classifier. We compare our proposed

TABLE I: Summary of the dataset

Training Images for Human Detection		
Images with human	**Images without human**	**Total**
354	600	954

Testing Image Set			
With-helmet	**Without-helmet**	**Non-human**	**Total**
67	83	50	200

TABLE II: Performance measures in helmet detection

Measure	Description
TP	Num. of images correctly classified as including worker (or with helmet)
TN	Num. of images correctly classified as excluding worker (or without helmet)
FP	Num. of images mistakenly classified as including worker (or with helmet)
FN	Num. of images mistakenly classified as excluding worker (or without helmet)
TPR	$TP/(TP+FN)$
FPR	$FP/(FP+TN)$
ACC	$(TP+TN)/(TP+TN+FP+FN)$

method for human (i.e., construction worker) detection with the method using HOG only. Based on the training data set with 354 human and 600 non-human image segments, we conduct 10-folds cross validation for evaluation. The results shown in Table III and Figure 8 indicate that extracting HOG features from DCT coefficients of the image is more effective in human (i.e., construction worker) detection than using HOG only.

TABLE III: Comparisons of different human detection methods

Method	**ACC (%)**	**TPR (%)**	**FPR (%)**
HOG (baseline)	81.13	74.59	15.83
DCT+HOG (proposed)	**91.93**	**80.01**	**3.33**

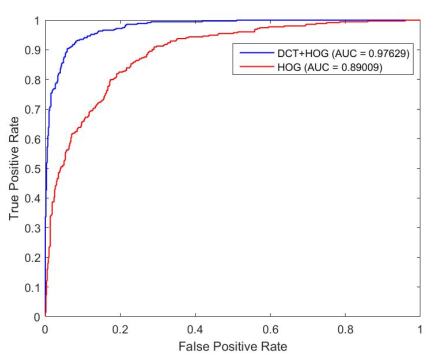


Fig. 8: ROC curves of different human detection methods

C. Evaluation of Helmet Use Detection

In this set of experiments, we further evaluate the performance of the proposed helmet use detection method. After the detection of object of interest (i.e., construction worker), the combination of color-based and Circle Hough Transform (CHT) feature extraction techniques is applied to detect helmet uses for the construction worker. Based on the 200 testing images (67 of which are tagged as "with-helmet", 83 are "without-helmet" and 50 are tagged as "non-human"), we compare our proposed method with the method using CHT only. The experimental results shown in Table IV and V demonstrate that combination of color-based and CHT feature extraction techniques outperforms using CHT only in helmet use detection. As same human detection algorithm is used in both cases, accuracy rates in detecting human and non-human are same. After identifying human objects (i.e., construction workers), as shown in Table V, in detecting helmet, the proposed method gives better accuracy (79.1%) than the baseline (67.16%); moreover, the proposed method is more successful in detecting the case of construction worker without helmet with 84.34% accuracy, while for baseline method it is only 45.78%. CHT tries to find all possible circles in the image, while inclusion of color information increases the accuracy of detection of the presence of the helmet. Without the color information, it fails to distinguish between circular helmet and human head as circular shape. That explains the reason behind the huge difference in detecting the case of construction worker without helmet.

TABLE IV: Comparisons of different helmet use detection methods

Method	**ACC (%)**
CHT (baseline)	61.0
Color + CHT (proposed)	**81.0**

TABLE V: Confusion matrix for different helmet use detection methods

CHT (Baseline)			
	Human	**Human**	**Non-human**
With-helmet	**67.16**	29.85	2.99
Without-helmet	54.22	**45.78**	0.00
Non-human	12.00	10.00	**78.00**
Color + CHT (Proposed method)			
	Human	**Human**	**Non-human**
With-helmet	**79.10**	17.91	2.99
Without-helmet	15.66	**84.34**	0.00
Non-human	12.00	10.00	**78.00**

VI. Conclusion and Future Work

In this paper, a novel approach is proposed for automatic detection of helmet uses for construction safety using computer vision and machine learning techniques. The proposed system has two major parts: one part incorporates frequency domain information of the image with a popular human detection algorithm HOG for human (i.e., construction worker) detection; the other part works for helmet use detection combining color information and Circle Hough Transform (CHT).

Currently, our system can detect helmets composed of some particular colors, such as yellow, blue, red, and white. As an extension of this work, we aim to make the system scalable to detect helmets with other colors. In future, the system will be made well capable of differentiating between normal cap and helmet, as the proposed system shows low performance in this case. Also, we aim to apply some deep learning techniques for improving the overall accuracy of the system. Also, applying upper body searching algorithm instead of detecting whole human as object of interest can improve the helmet detection accuracy.

Acknowledgement: The authors would like to acknowledge the contributions of the National Institute for Occupational Safety and Health (NIOSH) in providing constructive comments at various stages of this study.

Disclaimers: The findings and conclusions in the report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health (NIOSH). Mention of company names or products does not imply endorsement by NIOSH.

References

[1] Y.-S. Ahn, J. F. Bena, and A. J. Bailer. Comparison of unintentional fatal occupational injuries in the Republic of Korea and the United States. In Injury Prevention, Vol. 10(4), 199-205, 2004.

[2] Bureau of Labor Statistics. Safety and healthdangerous jobs. Compensation and Working Conditions. In http://www.bls.gov/iif/oshwc/cfar0020.pdf, 1997.

[3] Can Stock Photo. Roofing stock photos and images. In http://www.canstockphoto.com/images-photos/roofing.html

[4] Can Stock Photo. Roofing stock photos and images. In http://www.canstockphoto.com/images-photos/roofing.html#start:75

[5] Can Stock Photo. Roofing stock photos and images. In http://www.canstockphoto.com/images-photos/roofing.html#rows:200

[6] Center for Construction Research and Training. The Construction Chart Book: The U.S. Construction Industry and Its Workers. 5th edition ed. 2013, Silver Spring, MD: CPWR.

[7] N. Dalal, and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.

[8] DeBruyne, N.F. and A. Leland. American War and Military Operations Casualties: Lists and Statistics 2015. In Congressional Research Service, 2015.

[9] L. Ding, and A. Goshtasby. On the Canny edge detector. In Pattern Recognition, Vol. 34(3), 721–725, 2001.

[10] S. Du, M. Shehata, and W. Badawy. Hard hat detection in video sequences based on face features, motion and color information. In 3rd International Conference on Computer Research and Development (ICCRD), Vol. 4, 2011.

[11] J. Gavin. 2015 Construction Outlook: An Economic Recovery Finds Its Footing, 2015.

[12] D. Gavrila. Pedestrian detection from a moving vehicle. In ECCV, Vol. 2:37-49, 2000.

[13] Getty Images. Construction images. In http://www.gettyimages.com/search/more-like-this/91988434?assettype=image&excludenudity=true&family=creative&license=rf&sort=best

[14] Getty Images. Construction images. In http://www.gettyimages.com/search/more-like-this/91988434?assettype=image&excludenudity=true&family=creative&license=rf&page=2&sort=best

[15] Getty Images. Roof construction pictures and images. In http://www.gettyimages.com/photos/achillestarace?excludenudity=true&mediatype=photography&page=1&phrase=roof%20construction&sort=mostpopular

[16] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In ICDSC, 2008.

[17] Jaselskis, E.J., S.D. Anderson, and J.S. Russell. Strategies for achieving excellence in construction safety performance. In Journal of Construction Engineering and Management, 122(1): 6170, 1996.

[18] J. Han, M. Kamber, and J. Pei. Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2011.

[19] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in color images. In IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24(5), 696-706, 2002.

[20] R. Lienhart, and J. Maydt. An extended set of Haar-like features for rapid object detection. In Proc. of IEEE International Conference of Image Processing, Vol. 1, 900-903, 2002.

[21] A. Micilotta, E. Ong, and R. Bowden. Detection and Tracking of Humans by Probabilistic Body Part Assembly. In Proceedings on British Machine Vision Conference, Vol. 1, 429-438, 2005.

[22] K. Mikolajczyk, C. Schmid, and A. Zisserman. Human detection based on a probabilistic assembly of robust part detectors. In ECCV, Vol. 1, 69-81, 2004.

[23] D. T. Nguyen, W. Li, and P. O. Ogunbona. Human detection from images and videos: A survey. In Pattern Recognition, Vol. 51, 148-175, 2016.

[24] M. Nixon, and A. Aguado. Feature Extraction and Image Processing, 1st Edition. In Elsevier, Burlington, MA, USA, 2002.

[25] Occupational Safety & Health Administration. Constructions Fatal Four. In https://www.osha.gov/oshstats/commonstats.html, 2013.

[26] M.-W. Park, E. Palinginis, and I. Brilakis. Detection of construction workers in video frames for automatic initialization of vision trackers. In Construction Research Congress, 2012.

[27] Public Citizen. The Price of Inaction: A Comprehensive Look at the Costs of Injuries and Fatalities in Maryland's Construction Industry, 2012.

[28] D. Ramanan, D. Forsyth, and A. Zisserman. Strike a pose: tracking people by finding stylized poses. In CVPR, 1, 2005.

[29] T. Roberts, S. McKenna, and I. Ricketts. Human pose estimation using learnt probabilistic region similarities and partial configurations. In ECCV, Vol. 4, 291-303, 2004.

[30] M. D. Rodriguez, and M. Shah. Detecting and segmenting humans in crowded scenes. In Proceedings of the 15th international conference on Multimedia, 2007.

[31] R. Ronfard, C. Schmid, and B. Triggs. Learning to parse pictures of people. In ECCV, Vol. 4, 700-714, 2002.

[32] K. Sabottka, and I. Pitas. Segmentation and tracking of faces in color images. In Proc. of International Conference on Automatic Face and Gesture Recognition, 236-241, 1996.

[33] Shutter Stock. Roofing stock photos, illustrations, and vector rrt. In http://www.shutterstock.com/s/roofing/search.html?page=1&thumb_size=mosaic

[34] Shutter Stock. Roofing stock photos, illustrations, and vector rrt. In http://www.shutterstock.com/s/roofing/search.html?page=2&thumb_size=mosaic

[35] U.S. Bureau of Labor Statistics. National Census of Fatal Occupational Injuries in 2014 (Preliminary Results).

[36] P. Viola, M. Jones, and D. Snow. Detecting Pedestrians Using Patterns of Motion and Appearance. In IJCV, Vol. 63(2), 153-161, 2005.

[37] P. Viola, and M. Jones. Rapid object detection using a boosted cascade of simple features. In CVPR, Vol. 1, 511-518, 2001.

[38] C.-Y. Wen, S.-H. Chiu, J.-J. Liaw, and C.-P. Lu. The safety helmet detection for ATM's surveillance system via the modified Hough transform. In Proceedings of IEEE 37th Annual 2003 International Carnahan Conference on Security Technology, 2003.

[39] Y. Wu, and T. Yu. A Field Model for Human Detection and Tracking. In CVPR, 28, 2006.

[40] H. K. Yuen, and J. Princen, and J. Illingworth, and J. Kittler. Comparative study of Hough transform methods for circle finding. In Image and vision computing Vol. 8(1), 71-77, 1990.

[41] L. Zhao, and C. Thorpe. Stereo-and neural network-based pedestrian detection. In IEEE Transactions on ITS, Vol. 1(3), 148-154, 2000.

[42] L. Zhao, and L. Davis. Closely Coupled Object Detection and Segmentation. In ICCV, Vol. 1, 2005.