Vision-Based Safety Vest Detection in a Construction Scene

H. Seong, H. Choi, H. Cho, S. Lee, H. Son, and C. Kim*

Department of Architectural Engineering, Chung-Ang University, Seoul 06974, South Korea
E-mail: gusdn7543@cau.ac.kr, vinaj516@gmail.com, hyukmanjo@gmail.com, leesungwook@cau.ac.kr, hjson0908@cau.ac.kr, changwan@cau.ac.kr (* Correponding author)

Abstract

The computer vision-based detection of construction workers in images or videos is necessary for the safety managements and productivity of construction workers. Researchers in previous studies, detecting construction workers in the construction scene via computer vision techniques, have considered various features such as motion, shape, and color. Due to the pose changes of the workers, construction worker detection using body shape as a feature in the construction scene remains a challenging task. This study proposes a safety vest detection method, as a preceding method of the construction worker detection, which uses the motion of workers and the color pixels of safety vests for distinguishing from others in the construction scene independent of the workers' pose changes. The background subtraction method is performed by using an approximate median filter for the purpose of reducing the candidate regions that the color pixel classification will be performed as a sequential step. Then, the color pixel classification method is performed with a comparative analysis of two color spaces (Lab and HSV [hue, saturation, and value]) and three types of classifiers (a support vector machine [SVM], an artificial neural network [ANN], and a logistic regression [LR]) to classify the pixels of safety vests accurately. The proposed method has been tested on the actual construction site video streams. The proposed method of this study, thus far, is making progress toward the achievement of the robust and effective construction worker detection in the construction scene.

Keywords – Computer Vision; Sensing; Data Mining; Machine Learning; Detection; Classification; Image Processing; Construction Worker; Color Space Transformation

1 Introduction

The computer vision-based detection of construction workers is an important element for various applications in both safety management and productivity measurement. Computer vision provides a rich set of data about a construction scene that helps distinguish between workers and others that can be found on a construction site. Within the last decades, video surveillance has become common on construction sites with the deployment of vision sensors. The better the quality of the camera is, the higher the quality of the acquired images will be. Computer vision techniques, in combination with computer networks and automated image recognition, take advantage of this. The valuable information in video frames of complex construction tasks can be provided promptly and accurately with highly sophisticated computer vision techniques.

There have been a few studies on the use of computer vision techniques to detect workers in a construction scene [2–6]. The researchers in most previous studies employed the shape of the human body as a feature for construction worker detection. However, these studies were limited to detecting workers with upright postures. Although these methods may work in controlled construction scenarios, construction worker detection in construction scenes remains challenging because of the various postures of construction workers in construction scene. To meet these challenges, color-based features have obvious advantages over other features (especially in complex environments) because the color of an object of interest is independent from the positions and shapes [7]. Color-based features are commonly used in various object detection techniques such as the detection of workers, equipment, and materials in the construction sites [5,8,9]. The color property of the safety vest is expected to be an accurate and robust feature that has the potential to overcome the limitations of construction worker detection that uses shape-based features.

The American National Standards Institute [10] stipulated that wearing a high-visibility safety vest is mandatory for all construction workers for reducing the likelihood of accidents. Accordingly, the wearing of a safety vest identifies workers on a construction site.

Besides that, the American National Standards Institute has provided a specification for the colors of the safety vest: fluorescent yellow-green and fluorescent orange-red. Both the fluorescent yellow-green and fluorescent orange-red colors are not prevalently found in the construction scene; thus, they are distinctive colors that can be readily distinguished from the color of other objects in the construction scene.

This study proposes a safety vest detection method using the motion of the workers with the colors of safety vests being used in a preceding process of effective and reliable construction worker detection via a computer vision-based technique. The background subtraction method applying an approximate median filter was used for reducing the candidate region for the sequential step (i.e., color pixel classification). A data set of safety vest pixels was generated for comprehensive color pixel classification results. To achieve robust color pixel classification against the illumination changes, a transformation from RGB (red, green, and blue) to HSV and Lab color spaces (with or without illuminant components) was performed. A comparison of the performances among three classifiers (i.e., SVM, ANN and LR) with different color spaces was performed to find the most appropriate combination for the classification of safety vest pixels. The proposed method had been tested on the video streams of actual construction sites.

2 Literature Review

With computer vision, construction worker classification using color-based features is used as a first step in construction worker tracking and monitoring techniques on a frame-by-frame basis. Recently, a few researchers have applied the color-based features of safety vests to construction worker classification methods. Their approaches to classifying workers in frames are grouped into two types: a color histogram-based method and a pixel-based method.

A color histogram is widely used as one of the most common color-based features, due to its invariance of small changes in viewing position and its computational efficiency. Construction worker classification methods using color histograms have been introduced in previous studies. Park and Brilakis [2] proposed a method for the detection of construction workers when initializing a scene. They used the hue-saturation color property of the upper-body section, which involved using the safety vest to construct a histogram. Then, the histogram was classified into the categories of construction worker and non-construction worker by using the k-nearest neighbor (KNN) classification method. Memazardeh et al. [3] suggested using histograms of oriented gradients and hue-saturation color-based methods for the identification of workers on a construction site. The researchers used the color histogram of a region of interest to discriminate between construction workers and the background. In their approach, the support vector machine (SVM) classifier, which is one of the most powerful machine-learning techniques for solving binary classification problems on a large scale, was employed as a color-based classification method for construction workers. However, a color histogram does not include spatial information, especially for a large database [11]. Moreover, construction worker classification using a color histogram of the region of interest can disturb accurate performance results because the color histogram is affected by other surrounding colors in the region (e.g., colors of hardhats, clothes, and surroundings).

The state-of-the-art method of the pixel-based classification approach is the classification of individual pixels in the region of interest to safety vest and non-safety vest categories on the basis of pixel color properties. In the past few years, Gong and Caldas [4] proposed a construction worker detection method for the purpose of measuring productivity by using color pixel classification to recognize construction workers. Their approach used each pixel of safety vests as a feature. Their color-based model used the expectation by maximization method to estimate the parameters of the Gaussian models, which can determine the pixels of safety vests by clustering of the pixels. Their result presented 85.50 % for true positives and 5.39 % for false positives. The classification method of pixels-to-safety vest pixels using a Gaussian model may not produce optimal results in a construction scene in which the inherent color property of the safety vest is affected by variations such as illumination changes and the position and characteristics of the camera. Empirical studies have found that the ANN and SVM classifiers outperformed the Gaussian model in the object detection in the construction scene [12].

3 Methodology

This study proposes a method to reliably detect safety vests in the construction scene via computer vision techniques. The proposed method includes the background subtraction method and color pixel classification, which are separately performed in sequential steps. As the first step, the foreground regions of interest, which are in motion, are detected using the background subtraction method. The background subtraction method determines which pixels belong in the background and which pixels belong in the foreground region of interest. Then, the color pixel classification is performed to determine which pixels belong to safety vests.

3.1 Background Subtraction Method

Background subtraction is a commonly used detection method for identifying foreground objects in motion in the video stream. Because the method is desirable for achieving a high level of precision in the detection of foreground objects of interest, it is employed as a reliable first step in a number of computer vision applications. In this study, the background subtraction method was performed for the purpose of reducing the areas to be computed. This study selected the background subtraction method using the approximate median filter, which was introduced by McFarlane and Schofield [13], among many highly sophisticated background subtraction algorithms. In the field of image processing, the approximate median filter has been commonly used for removing impulsive noises because of the simplicity of its process in that it is computationally efficient. In construction video surveillance applications that work with outdoor scenes, the background scene includes a number of nonstationary objects such as the construction materials, tree branches, and soil with which movement depends on the wind. These nonstationary objects cause an increase in computational burdens due to false-positive detection. Therefore, this study eliminated such foreground regions by using the approximate median filter. The method of using the approximate median filter has been validated as a reliable method when applied to construction videos for the purpose of distinguishing the foreground regions from the background scene [2].

3.2 Color Pixel Classification Method

For the color pixel classification (the detection considered in this study) the objective was to determine whether each pixel belongs to a safety vest color or non-safety vest color by using only a color triplet as input. The major difficulty in color pixel classification in the construction scene is that the color property of the safety vest is affected by variations in the intensity of illumination according to environmental factors (e.g., weather condition, time of the day, and seasonal variations). For a reliable classification method, the approach has to be robust against such variations.

3.2.1 Data Collection and Data Pre-Processing

To achieve a reliable detection result, a comprehensive collection of data is essential. This study included the use of more than 15,000 images that were extracted from 52 different video sequences. The video sequences were acquired from actual construction sites, for a data set considering such variations. The videos were selected in consideration of different conditions such as the viewing position of the camera, the time of day (between 7am and 5pm), and the background scene. Each image in the data set included at least one construction worker who was wearing a safety vest as shown in Fig. 1. Each image was divided into the small sub-images with a size of 30×30 pixels. Then, the sub-images were categorized and labeled as "fluorescent yellow-green," "fluorescent orange-red," "background," or "indeterminate." Next, the data set from the sub-images that were categorized as "indeterminate" were excluded from the evaluation because that data set could not be defined as either the pixels of safety vest colors or the background pixels. The background images, which are of non-safety vest regions, include all kinds of scenery such as construction equipment, sky, soil, trees, construction materials, clothes of workers, and other construction-related objects. The data collected from safety vest color and non-safety vest color sub-images amounted to approximately 45,000 pixels each for "fluorescent yellow-green" and "fluorescent orange-red," which are safety vest colors, and 90,000 pixels for the background.

Typically a construction worker detection framework using color-based features includes the transformation of RGB into other color spaces, because RGB color spaces are subjected to the deterioration of color property in variations of illumination. It is generally assumed that variations in color property occur more in chrominance components, meaning that excluding the illuminance component and using chrominance components only leads to a better performance. Therefore, this study considered the two types of color spaces: Lab and HSV. These color spaces are the general choices among color spaces in construction worker detection using color features [2–4]. Moreover, this study included an investigation into the role of illuminance components by excluding illuminance components: the L of Lab and the V of HSV.

3.2.2 Measuring Safety Vest Color Models

This study considered three different types of classifiers: SVM, ANN, and LR. These classifiers are generally used in color pixel classification methods. The algorithms of WEKA release 3.8.1, a typical Java-based machine learning program, were used in this study. Defined by its architecture, ANN consists of one or more hidden layers of neurons. The connections between them are indicated by weights. In this study, we used a back-propagation neural network. The SVM has been prevalently used for solving a large number of complex binary classification problems [14]. The SVM model learns the high-dimensional data, and implements generalization performance well, with the aid of multiplier parameters like the Lagrange multiplier. The LR classifier is a general classifier of linear regressions

[15]. An advantage of LR is that it can generate a simple probabilistic formula for the classification. A set of parameters was adjusted for each classifier. For assessing the performance of the classifiers with the transformed color spaces, a k-fold cross-validation was used in this study. It has been determined that a 10-fold cross-validation is an optimal choice in terms of error estimation and computation; therefore, this study employed a 10-fold cross-validation to assess the performance of each combination.

In this study, the performance was evaluated by two measures: precision and recall. These measures, which stand in opposition to one another, are widely used in the evaluation of both color pixel classification and construction worker detection. Precision (see Equation (1)) and recall (see Equation (2)) are defined as follows, where the true positive (TP) is the number of safety vest pixels that are correctly classified as safety vest pixels, the false negative (FN) is the number of safety vest pixels that are incorrectly classified as background pixels, and the false positive (FP) is the background pixels that are incorrectly classified as safety vest pixels. Tables 1 and 2 seperately present the classification results of each combination for the two colors of safety vests. The results potentially imply that the color pixel classification of safety vests can be achieved by using color pixel information alone, regardless of the type of color space and classifier.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

Table 1. Classification results of fluorescent yellow-green color pixels.

Color space	Classifier	Precision (%)	Recall (%)
Lab	ANN	98.30	**98.48**
	SVM	**99.36**	97.05
	LR	95.27	95.59
ab	ANN	97.12	97.65
	SVM	98.33	98.16
	LR	94.78	94.04
HSV	ANN	97.76	97.64
	SVM	88.27	93.75
	LR	72.40	77.46
HS	ANN	93.24	96.62
	SVM	94.84	86.67
	LR	91.59	95.31

Table 2. Classification results of fluorescent orange-red color pixels.

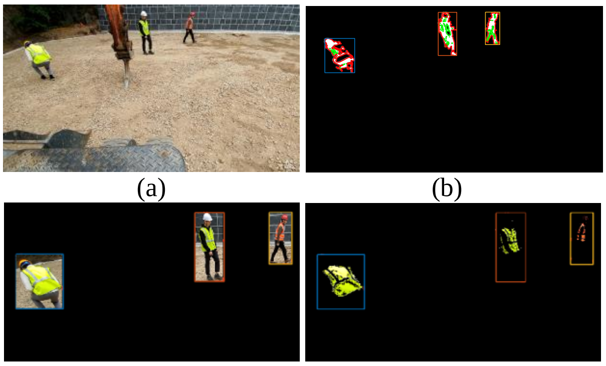
Color space	Classifier	Precision (%)	Recall (%)
Lab	ANN	86.43	93.99
	SVM	**94.94**	**95.15**
	LR	87.61	88.92
ab	ANN	86.09	91.01
	SVM	87.47	92.10
	LR	86.34	88.43
HSV	ANN	87.34	94.30
	SVM	88.25	93.75
	LR	82.26	86.71
HS	ANN	79.25	87.35
	SVM	78.92	88.28
	LR	72.91	77.84

Table 1 and 2 present the classification results of each combination for two colors of safety vests separately. The results potentially imply that the color pixel classification of safety vest can be achieved by using color-pixel information alone, regardless of the types of color space and classifier. As for the overall performance of the combinations of color spaces and classifiers, the SVM classifier with the Lab color space demonstrated the best performance on the data sets of fluorescent yellow-green in terms of precision, and fluorescent orange-red in terms of both precision and recall, except in terms of recall on the data set of fluorescent yellow-green, in which the ANN classifier fared the best. Although the ANN classifier with the Lab color space yielded better results only in terms of recall on the data set of fluorescent yellow-green, the difference in performance between the SVM classifier and the ANN classifier with the Lab color space were significant on the fluorescent orange-red in both measures.

4 Results

To assess the effectiveness of the proposed methods, each method was tested on the 58 image sequences that were processed and collected from the 12 videos of an actual construction site. Figure 1 shows the color pixel classification results on an image captured from actual construction site video. Construction workers wearing safety vests with different postures (both upright and squatting postures) are in the image. The background subtraction method using the approximate median filter was applied as a first step for detecting foreground objects of interest. Figure 1(b) shows that the proposed background subtraction method correctly detects the foreground regions that are potential candidates for the color pixel classification method as sequential step without any impulsive noises. The detected foreground regions are enlarged for the purpose of a better display, as shown in Figure 1(c).

After the background subtraction was performed, each pixel of the detected foreground regions with bounding box were classified as either pixels of safety vests or pixels of non-safety vests. The color pixel classification was performed by the SVM classifier with color space transformation into the Lab color space, which presented the most desirable performance. The pixels of the safety vest that were detected by the proposed method appeared as color pixels, and the pixels of the non-safety vest are black as shown in Figure 1 (d). These results indicate that the proposed method is useful and reliable. Despite such variations resulting from illumination change, most of the safety vest pixels were correctly detected as belonging to safety vests. These results indicate that the proposed method is invariant against changes in workers' postures and illuminant variation has a reliable detection accuracy.



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

Figure 1. Color pixel classification results on the actual construction scene (a) original image; (b) foreground region detection results by background subtraction method; (c) enlarged image of foreground region detection result (depicted as a color image); and (d) color pixel classification results of safety vest pixels.

5 Conclusion and Recommendation

To achieve the detection of construction workers in various postures, this study proposed a safety vest detection method of using both motion of the construction workers and the colors of safety vests. The proposed method involved a common background subtraction method using the approximate median filter to detect the foreground region to reduce the candidate regions. As a sequential step, each pixel in the regions was classified as being safety vest pixels or non-safety vest pixels. To demonstrate which combination of color space and classifier performs the best, a comparative analysis was conducted. The assessment was carried out on the performance of two color spaces with three classifiers. The results showed that the combination of the Lab color space and the SVM classifier had the most desirable outcomes. The proposed method was tested on the actual construction site video, and the results indicated that safety vest detection was achieved with a high-level of accuracy for both the upright and squatting postures of construction workers.

The proposed method of this study, thus far, is expected to make progress toward the achievement of the robust and effective construction worker detection in the construction scene. Future work will include the method that specifically identifies the construction workers. In addition, the method used in this study can only be applied using a static camera, which is a common limitation of the background subtraction method; therefore the greatly developed method will be investigated.

References

[1] Seo, J. S., Han, S. U., Lee, S. H., and Kim, H. K. Computer vision techniques for construction safety and health monitoring. Advanced Engineering Informatics, 29(2): 239–251, 2015.

[2] Park, M. W. and Brilakis, I. Construction worker detection in video frames for initializing vision trackers. Automation in Construction, 28: 15–25, 2012.

[3] Memarzadeh, M., Golparvar-Fard, M., and Niebles, J. C. 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, 2013.

[4] Gong, J. and Caldas, C. H. An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations. Automation in Construction, 20(8): 1211–1226, 2011.

[5] Chi, S. and Caldas, C. H. Automated object identification using optical video cameras on construction sites. Computer-Aided Civil and Infrastructure Engineering, 26(5): 368–380, 2011.

[6] Yang, J., Arif, O., Vela, P. A., Teizer, J., and Shi, Z. Tracking multiple workers on construction sites using video cameras. Advanced Engineering Informatics, 24(4): 428–434, 2016.

[7] Son, H., Kim, C., Hwang, N., Kim, C., and Kang, Y. Classification of major construction materials in construction environments using ensemble classifiers. Advanced Engineering Informatics, 28(1): 1–10, 2014.

[8] Neto, J. A., Arditi, D., and Evens, M.W. Using colors to detect structural components in digital

pictures. Computer-Aided Civil and Infrastructure Engineering, 17: 61–67, 2002.
[9] Zou, J. and Kim, H. Using hue, saturation, and value color space for hydraulic excavator idle time analysis. Journal of Computing in Civil Engineering, 21: 238–246, 2007.
[10] ANSI/International Safety Equipment Association (ISEA), American national standard for high-visibility safety apparel and headwear, ANSI/ISEA 107–2010, Washington, DC, 2010.
[11] Huang, J., Kumar, S. R., Mitra, M., Zhu, W. J., and Zabih, R. Image indexing using color correlograms. In Procedings of Computer Vision and Pattern Recognition, page 762–768, 1997.
[12] Son, H., Kim, C., and Kim, C. Automated color model–based concrete detection in construction site images by using machine learning algorithms. Journal of Computing in Civil Engineering, 26(3): 421–433, 2011.
[13] McFarlane, N. J. and Schofield, C. P. Segmentation and tracking of piglets in images. Machine vision and applications, 8(3): 187–193, 1995.
[14] Cortes, C. and Vapnik, V. Support-vector networks. Machine learning, 20(3): 273–297, 1995.
[15] Hastie, T., Tibshirani, R., and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, 2001.