

A Comparative Study of Machine Learning Classification for Color-based Safety Vest Detection on Construction-Site Images

Hyeonwoo Seong*, Hyojoo Son**, and Changwan Kim***

Received November 6, 2017/Revised January 30, 2018/Accepted March 4, 2018/Published Online September 28, 2018

Abstract

Detecting the safety vests is an important foundation for various applications in safety management and productivity measurement. The fluorescent yellow-green color and fluorescent orange-red color of safety vests are generally considered as the most distinctive colors which represent workers in construction-site images. The objective of this study is to provide an evaluation of the safety vest detection using color information in construction-site images. The data sets of two colors of safety vests and the background were generated and used in this study. A comparative analysis of combinations of five color spaces (RGB, nRGB, HSV, Lab, and YCbCr) and six classifiers (ANN, C4.5, KNN, LR, NB, and SVM) was conducted. The performance of each combination was assessed in terms of the precision, recall, and F-measure. Moreover, an evaluation of the effects of color space conversion and the absence of luminance components on the detection performance was conducted. The comparison results showed that C4.5 classifier combined with YCbCr and SVM classifier combined with Lab, respectively, outperformed other combinations on each data set of safety vest colors. Furthermore, RGB color space transformation into non-RGB color spaces enhanced the classification performance. The evaluation also showed that the removal of luminance components did not help to improve the performance.

Keywords: color space transformation, data mining techniques, image processing, machine learning, pixel-level classification, safety vest detection, worker detection

1. Introduction

The colors of safety vests are generally considered as the most distinctive colors of the construction workers in computer vision (Seo *et al.*, 2015). The American National Standards Institute (ANSI, 2015) stipulates that wearing high-visibility safety vests is mandatory for construction workers to reduce the likelihood of accidents. Furthermore, the American National Standards Institute provides specifications for the two colors of safety vests: fluorescent yellow-green and fluorescent orange-red. Fluorescent yellow-green and fluorescent orange-red are not only rarely found on other objects at construction sites, but also they are distinctive and synthetic colors that can easily be distinguished from other colors on construction-site images. Detecting safety vests using their color information thus helps with distinguishing construction workers from others on construction-site images. For the process of color-based safety vest detection considered in this study is to distinguish individual pixels of safety vest colors from pixels of backgrounds.

A key challenge in detecting safety vests using color information is that color information may be affected by environmental conditions such as lighting condition changes, the weather, the background color, and characteristics of the camera in computer vision

(Kakumanu *et al.*, 2007). These variations cause performance degradation in detecting safety vests. In response to these challenges, color space transformation is typically performed to reduce the distortions of color information in various applications of color-based object detection approaches (e.g., construction materials Son *et al.*, 2014 and equipment Zou and Kim, 2007). Typically, red, green, and blue (RGB) components are the primary colors processed on cameras and computers (Platanotis and Venetsanopoulos, 2013). Because RGB is sensitive to changes in illumination (Mohamed *et al.*, 2007), RGB is converted to various color spaces to gain robustness against illumination changes. However, no color space is a dominant choice because color can be represented in different ways (Nallaperumal, 2007). Furthermore, detection performance can be improved by selecting the most optimal classifier in machine learning considering representation of data. Therefore, challenges with the detection of safety vests on construction-site images can be defined as follows: the representation of the pixels of safety vests with the most appropriate color space, and the modeling of safety vest and non-safety vest color pixels with the best-performing classifier.

Recently, a few studies have used the colors of safety vests as a feature for detecting construction workers in images and videos. The color distribution models of their color approaches are

*Member, Research Assistant, Dept. of Architectural Engineering, Chung-Ang University, Seoul 06974, Korea (E-mail: gusdn7543@cau.ac.kr)

**Research Assistant, Dept. of Architectural Engineering, Chung-Ang University, Seoul 06974, Korea (E-mail: hjson0908@cau.ac.kr)

***Member, Professor, Dept. of Architectural Engineering, Chung-Ang University, Seoul 06974, Korea (Corresponding Author, E-mail: changwan@cau.ac.kr)

grouped into two types: a color histogram model and per-pixel model. Park and Brilakis (2012) proposed a method for the detection of construction workers when initializing a scene. They applied methods that used motion and shape as features to reduce candidates for the color histogram method as a sequential step. The upper-body section of a construction worker was used to construct a color histogram. They extracted the hue and saturation components from the Hue, Saturation, and Value (HSV) color space. Then, each color histogram was classified into the two categories of construction worker (a person wearing a safety vest) and non-construction worker (a person not wearing a safety vest) by using the K-Nearest Neighbor (KNN) classifier. In the study of Memazardeh *et al.* (2013), they presented a color model-based detection method for construction workers and equipment. They used histograms of gradients and colors which were generated from hue and saturation components of HSV color space as feature descriptors. Then identification of candidates for construction workers and equipment on images was facilitated by using a Support Vector Machine (SVM) classifier. However, the proposed method is considered to be unsuitable for real-time applications in terms of safety management due to its relatively high computational cost. Furthermore, their approach has a certain limitation in detecting construction workers with various working postures because their data sets of this study only include standing construction workers.

The state-of-the-art method of the per-pixel model approach is the classification of individual pixels into safety vest and non-safety vest categories on the basis of the color representation. Gong and Caldas (2011) proposed a detection method for assessing the productivity of construction workers and equipment by using color information. The conversion from RGB to Lab (L is a luminance component and the other two are chrominance components) color space was conducted. In their approach, only the chrominance components were used to build a model because they showed the most compact representation. They used the individual pixels of safety vests to build a Gaussian mixture model which can be used to identify the pixels of safety vests based on the clustering of the pixels. Their results presented 85.50% for the true positives and 5.39% for the false positives. However, using only chrominance components and excluding luminance components may not lead to the improvement of classification performance because luminance component has potential to provide more information for separating between safety vest and non-safety vest pixels. Khan *et al.* (2012) demonstrated the role of luminance component in skin detection using color-based methods. Furthermore, Son *et al.* (2012) found that Artificial Neural Network (ANN) and SVM classifiers outperformed the Gaussian mixture model in the color-based detection of concrete in construction-site images. The color spaces and classifiers which were employed in previous studies on construction worker detection by using color the information of safety vests may be applied and perform well only on the controlled and limited construction-sites.

The objective of this study is to provide a comprehensive

evaluation of safety vest detection using color information in construction-site images. The data sets of two colors of safety vests (i.e., fluorescent yellow-green and fluorescent orange-red) and the background (e.g., construction equipment, sky, soil, and trees) were generated and used in this study. To propose the best performing model, a comparative analysis was conducted with five color spaces (RGB, normalized RGB [nRGB], HSV, Lab, and YCbCr) and six classifiers (ANN, Commercial 4.5 [C4.5], KNN, Logistic Regression [LR], Naïve Bayes [NB], and SVM). Each model was assessed by using three measures (precision, recall, and F-measure). Moreover, the evaluation of the effects of a conversion from RGB to nRGB, HSV, Lab, and YCbCr spaces and the absence of luminance components (V of HSV, L of Lab, and Y of YCbCr) on the detection performance was conducted along with a comparative analysis.

2. Experimental Setup

2.1 Data Collection

To achieve a reliable result with a comparative analysis, a comprehensive collection of data is essential. The color property of a construction scene is affected by many factors such as variations in the illumination, background, characteristics of the image sensor, and its view point. Accordingly, to cover these variations, this study generated a comprehensive data set for the safety vest detection.

This study collected more than 15,000 images that have been extracted from 52 different video sequences. These video sequences were captured by CMOS image sensors. In recent years, the quality of CMOS image sensors has been improved remarkably (Jiang *et al.*, 2017). Today, CMOS image sensors can provide a relatively high image quality compared to CCD image sensors with low power consumption. With this improvement of technology, CMOS image sensors are used for a broad range of construction industry. Videos were selected with consideration of different conditions such as the viewing position of the camera, the time of



Fig. 1. An example of Construction Site Image for Safety Vest Detection

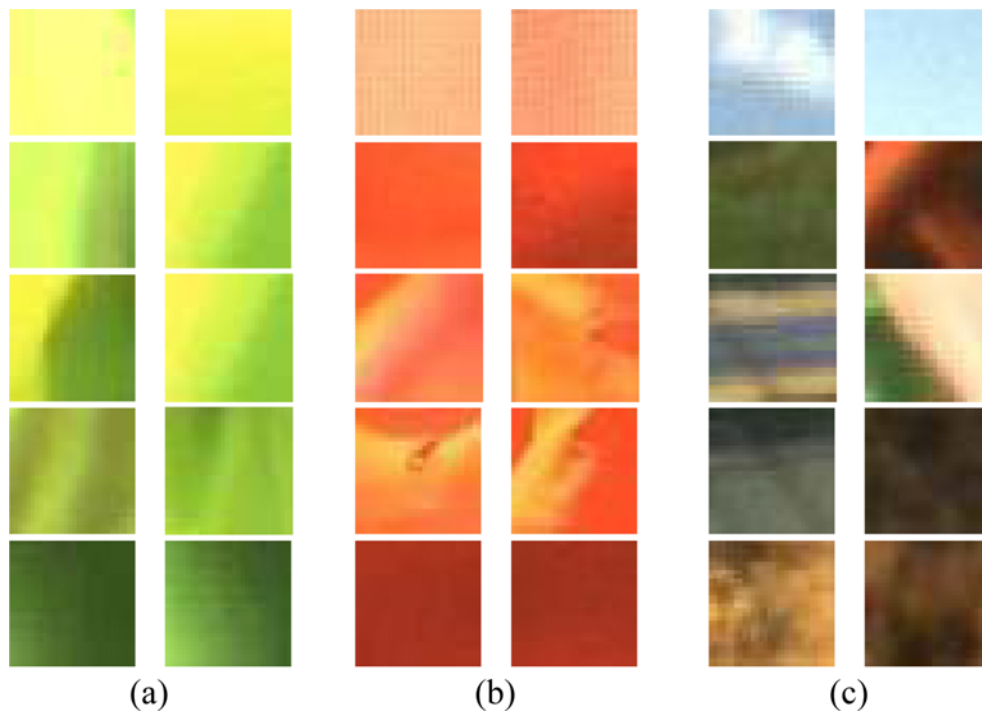


Fig. 2. (a) Examples of Fluorescent Yellow-green Color for Safety Vest Detection, (b) Examples of Fluorescent Orange-red Color for Safety Vest Detection, (c) Examples of Background Image (Non-safety vest colors)

the day (between 7am and 5pm), the weather, and the background sceneries.

The images, which extracted from the video sequences, of the data set included at least one construction worker who is wearing a safety vest as shown in Fig. 1. Each image was divided into sub-images with a size of 30×30 pixels. Then, the sub-images were both categorized and labeled as “fluorescent yellow-green,” “fluorescent orange-red,” “background,” or “indeterminate.” Next, the data set from the sub-images that are categorized as “indeterminate” were excluded from the evaluation. Because sub-images categorized as “indeterminate” could not be defined as either the pixels of safety vest colors or the pixels of non-safety vest colors. Fig. 2(a) and Fig. 2(b) show that the colors and textures of a safety vests varied from one to another, and present the included images from the data set under various conditions. Fig. 2(c) shows examples of background sub- image which are non-safety vest colors. These images include all kind of scenery such as construction equipment, sky, soil, trees, construction materials, clothes of workers, and other construction related objects. The data sets of safety vest colors and non-safety vest colors which generated from the images amounted to approximately 50,000 pixels for “fluorescent yellow-green” and “fluorescent orange-red” each (safety vest colors) and 50,000 pixels for the background (non-safety vest colors).

2.2 Data Pre-processing

For an effective color-based detection of safety vests, reducing the impact that variations will have on the color property of a safety vests is a critical issue. Although the colors of safety vests

are distinctive from others in a construction scene, a variation of a color property has occurred as shown in Fig. 2. Such a variation in color property is affected by changes in the illumination condition, and potentially degrades the performance in classifying color pixels.

In this study, investigation on the impact of a color space transformation on a color-based safety vest detection with five color spaces; RGB, nRGB, HSV, Lab, and YCbCr has conducted. This study uses equations of color space transformation as presented and applied in previous studies (Cheng *et al.*, 2001; Shaik *et al.*, 2015; Zou *et al.*, 2012). These color spaces involve both variance and invariance components to illumination variations. Moreover, to investigate the role of illumination components (V of HSV, L of Lab, and Y of YCbCr), a comparative analysis among color spaces with or without illumination components was conducted.

3. Methodology

3.1 Color Model of Safety Vests

This study used six different types of machine learning algorithms for the classification of safety vest pixels by color models of safety vests which are generated from the data sets. This includes an ANN, KNN, SVM, C4.5, NB, and LR. These classifiers are commonly used in classification methods (Witten *et al.*, 2016). The classification using the following algorithms are performed by WEKA 3.8.1, which is a Java-based machine learning tool. For each classifier, a set of parameters can be adjusted. During the study the default set came into practice. A

description of each classifier is given below.

3.1.1 Artificial Neural Network (ANN)

ANN is defined by its architecture which consists of one or more hidden layers of neurons. Connections between them are indicated by weights. In this study, we used a Back-Propagation (BP) neural network that is known as multilayer perceptron. This is a highly-sophisticated classification technique capable of modeling complicated non-linear functions (Delen *et al.*, 2005; Son *et al.*, 2014). The number of nodes for each hidden layer was set to 3. As an activation function, the sigmoid transfer function was selected. The learning rate that determines the amount of weights that are updated was set to the value of 0.3, and the momentum applied to the weights during updating was set to the value of 0.2.

3.1.2 K-nearest Neighbors (KNN)

The KNN is a non-parametric classifier, and an instance-based learning approach. It makes a decision for a class label of a new instance by major k-nearest instances in the training set using a distance metric (Holmes and Adams, 2002). The variable k represents the number of neighbors for each node (Derelioğlu, and Gürgeç, 2011). Usually, this is an odd number and is selected to reduce the occurrence of confusion between neighboring classes. As described by Duda *et al.* (2001) and El Gayer *et al.* (2006), the KNN classifier gains its robustness when a noisy training data set is used. Since KNN needs to compute the distance of every instance for all the training data, the high processing cost is considered as a major drawback of the KNN algorithm. This study makes use of the Euclidean distance as a distance metric, and the optimal value of k was set to 1.

3.1.3 Support Vector Machine (SVM)

The SVM has been prevalently used for solving a large number of complex binary classification problems (Chang and Lin, 2011; Kim *et al.*, 2011). The SVM can detect a hyperplane that maximizes the margin for inter-class separation (Dreiseitl *et al.*, 2001). Furthermore, the SVM uses a linear model to implement nonlinear class boundaries when the data cannot be separated linearly. It maps the original input feature vector spaces to a high-dimensional feature space by using a kernel function of the original data (Chen, 2011). The SVM model learns the high-dimensional data, and implements generalization performance well, with the aid from multiplier parameters like the Lagrange multiplier. In this study, as the kernel function we have used a radial basis function (RBF) (Ali and Smith, 2006) which is one of the most common choices of a kernel function. The value of C , which is the complexity parameter, and the tolerance parameter was set to 1 and 0.001, respectively.

3.1.4 Commercial 4.5 (C4.5)

C4.5 is an advanced version of the decision tree algorithm which builds trees for a rooted binary classification. It applies a set of rules using input feature vectors to divide a set of data into

smaller subgroups based on the concept of information entropy (Quinlan, 2014). There are three nodes for C4.5 which are the root node, the branch node, and the leaf node. Each node plays a key role in C4.5. The performance of the classification is influenced by both a confidence factor which is used for pruning, and the minimum number of instances per leaf. In this study a confidence factor and the minimum number of instances per leaf was set to 0.25 and 2, respectively.

3.1.5 Naïve Bayes (NB)

The NB classifier (Friedman *et al.*, 1997) is a simple linear probabilistic classifier based on the “Bayes theorem” with theoretically well-founded mathematical models to classify test instances with known instances. It calculates the maximum posterior probability for unknown instances. Then, the classifier assigns the most likely class to a given instance sample as represented by a feature vector. The NB classifier is widely used as a classification algorithm in many previous studies because it shows relatively accurate and efficient classification performance in high-dimensional data sets compared to decision tree or neural network-based algorithms (Rish, 2001).

3.1.6 Logistic Regression (LR)

The LR classifier is a general classifier of linear regression (Hosmer Jr. *et al.*, 2013). It can predict a discrete output from a set of training feature vectors which likely to be discrete, dichotomous, continuous, or concurrence of any of these. An advantage of LR is that it can generate a simple probabilistic formula for the classification. Furthermore, it does not require an assumption of linearity in the relationship between the training feature vectors, nor does it require a normal distribution of training feature vectors.

3.2 Evaluation

In this study, three performance measures were used for the evaluation of the classification results: precision, recall, and F-measure. Precision and recall are commonly used in information retrieval in the field of detection (Fritsch *et al.*, 2013; Lin and Yang, 2013; Satio and Rehmsmeier, 2015). Precision (see Eq. (1)) and recall (see Eq. (2)) are defined as follows: where the True Positive (TP) is the number of safety vest pixels correctly classified as safety vest pixels; the False Negative (FN) is the number of safety vest pixels incorrectly classified as background pixels; and the False Positive (FP) is background pixels incorrectly classified as safety vest pixels.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

Precision and recall stand in opposition to one other in that precision estimates insertion and substitution errors while recall estimates deletion errors. The F-measure has been defined as a

single performance measure that is a harmonic means of weighting precision and recall to estimates insertion, substitution, and deletion errors. The F-measure is often used to evaluate the classification performance of information retrieval (Power, 2011; McDonald and Price, 2012; Taha and Haunbury, 2015). The F-measure used in this study is given by the following formula:

$$F - \text{measure} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

For evaluating the classification performance of each combination of the six color spaces and six classifiers, k -fold cross validation was used. The k -fold cross validation method is reliable method to yield minimal variance and bias compared to all other validation methods (Hastie and Tibshirani, 2009; Arlot and Celisse, 2010). In k -fold cross validation, the data set is divided into k (positive integer) almost equal folds. Each fold in turn is then used for testing, and the remaining $k-1$ folds are used for training. The performance estimation of k -fold cross validation is calculated by simply averaging the individual performance measures. Most studies on large-scale data sets have determined that 10-fold cross validation is appropriate in terms of estimation and computation of error (Wei *et al.*, 2013). Accordingly, in this study, 10-fold cross validation was employed to assess the performance of each combination.

4. Results

In this section, the classification results obtained by the 10-fold cross validation were analyzed using three performance evaluation measures: precision, recall, and F-measure. The 10-fold cross validation was applied to the data sets of the fluorescent yellow-green color and fluorescent orange-red color separately. To examine the necessity of color space transformation and the role of luminance components of each color space, comparative analysis was carried out. Furthermore, McNemar's test was performed to assess whether or not the classification results were statistically significant.

Precision represents the percentage of color pixels correctly classified as safety vest pixels with respect to the total number of color pixels classified as safety vest pixels, and recall represents the percentage of correctly classified safety vest pixels with respect to the total number of actual safety vest pixels. Tables 1 and 2 show the precision and recall of the classification results generated by 10-fold cross validation. The results indicate that the performances of C4.5 and KNN classifiers are stable and performed relatively better than other classifiers on both data sets with both measures, ranging from 94.97 to 97.26 in terms of precision and from 94.74 to 98.95 in terms of recall for C4.5 and from 93.95 to 97.88 in terms of precision and 95.34 to 98.58 in terms of recall for KNN. The performances of the NB and LR were significantly worse than the performance of other classifiers in terms of both measures. These results lead to the conclusion that NB and LR classifiers are not optimal choices for classifying pixels of safety vests. A potential reason for the poor performance of

Table 1. Precision of Classification Performance on the Data Sets of: (a) The Fluorescent Yellow-green Color, (b) The Fluorescent Orange-red Color

(a)						
	ANN	C4.5	KNN	LR	NB	SVM
RGB	94.83	97.26	97.31	93.82	73.64	97.90
nRGB	91.71	94.63	95.36	90.87	91.30	90.48
HSV	92.66	97.23	97.24	73.23	79.07	94.22
Lab	94.99	97.26	97.31	94.39	95.50	97.51
YCbCr	94.67	97.32	97.42	93.83	91.74	97.56

(b)						
	ANN	C4.5	KNN	LR	NB	SVM
RGB	93.66	97.88	97.88	92.88	82.18	99.25
nRGB	90.42	95.19	93.95	91.29	91.18	91.73
HSV	94.32	98.06	97.75	90.62	76.30	95.10
Lab	93.56	97.88	97.90	92.24	89.25	98.56
YCbCr	92.71	97.97	97.84	92.88	86.19	98.69

Table 2. Recall of Classification Performance on the Data Sets of: (a) The Fluorescent Yellow-green Color, (b) The Fluorescent Orange-red Color

(a)						
	ANN	C4.5	KNN	LR	NB	SVM
RGB	95.70	98.78	98.58	92.17	77.40	92.82
nRGB	94.38	96.06	96.13	92.51	84.56	94.36
HSV	96.00	98.95	98.58	75.40	83.24	97.85
Lab	96.47	98.92	98.55	93.67	87.71	98.60
YCbCr	96.30	98.91	98.56	92.21	80.15	97.83

(b)						
	ANN	C4.5	KNN	LR	NB	SVM
RGB	96.27	98.48	98.55	89.10	77.79	94.38
nRGB	93.56	94.74	95.34	84.09	82.48	84.68
HSV	95.30	98.49	98.36	92.55	90.39	97.05
Lab	96.52	98.46	98.51	88.51	94.03	98.42
YCbCr	96.00	98.44	98.49	89.09	85.83	97.93

NB classifier is that the conditional independence assumption of NB classifier is rarely identical to the real-world. The combination of RGB color space and SVM classifier achieved the best performance in terms of precision. However, its performance in terms of recall was not as good as its performance in terms of precision. The combination of HSV color space with C4.5 classifier and RGB color space with KNN classifier achieved the best performance in terms of recall on fluorescent yellow-green and fluorescent orange-red data sets.

In order to propose the best performing combination for classification model on the data sets of safety vests' colors, Table 3 shows the F-measure and the results of McNemar's test of 10 combinations which arranged in descending order of F-measure. The McNemar's test is used to assess differences with statistical significance between classification performances of two combinations by testing the differences in frequencies of class

Table 3. F-measure of Classification Performance on the Data Sets for both the Fluorescent Yellow-green Color and the Fluorescent Orange-red Color. A p -value Represents a Result for McNemar's Test

Fluorescent yellow-green				Fluorescent orange-red			
Color space	Classifier	F-measure	p -value	Color space	Classifier	F-measure	p -value
YCbCr	C4.5	98.11	-	Lab	SVM	98.49	-
Lab	C4.5	98.08	0.481	YCbCr	SVM	98.31	0.000*
HSV	C4.5	98.08	0.234	HSV	C4.5	98.27	0.000*
Lab	SVM	98.05	0.000*	RGB	KNN	98.21	0.000*
RGB	C4.5	98.01	0.054	YCbCr	C4.5	98.20	0.000*
YCbCr	KNN	97.99	0.000*	Lab	KNN	98.20	0.000*
RGB	KNN	97.94	0.003*	RGB	C4.5	98.18	0.000*
Lab	KNN	97.93	0.001*	Lab	C4.5	98.17	0.000*
HSV	KNN	97.91	0.022*	YCbCr	KNN	98.16	0.000*
YCbCr	SVM	97.69	0.000*	HSV	KNN	98.05	0.000*

Note. An asterisk indicates statistical significance (corresponding $p < 0.05$).

labels for each instance (Tarabalka *et al.*, 2010; Petri *et al.*, 2012; Ciatto *et al.*, 2013; Radtke, 2015). Typically in area of statistics, the p -level to accept level of statistical significance is set at 0.05 (Rice, 1989, MacInnes *et al.*, 2016). In this study, if a p -value is less than 0.05, the performance of the combination which has the highest F-measure is better than that of the other combination with statistical significance. For the data set of fluorescent yellow-green color, the combination of YCbCr color space and C4.5 achieved the best performance in terms of F-measure. According to the result of the p -values, difference is not significant when compares to the Lab, HSV, and RGB color spaces combined with C4.5. Since C4.5 classifier can deal with both discrete and continuous attributes, the results of C4.5 have shown to be relatively stable for changes in attribute values (i.e., color space transformation). These results demonstrate that C4.5 classifier yields better performance than other classifiers regardless of the choice of the color space. Compared to other combinations which used SVM and KNN classifiers, the difference is significant that the p -values are lower than the acceptance level (corresponding $p < 0.05$). The best performance in terms of F-measure is achieved by the combination of Lab color space and SVM classifier on the data set of fluorescent orange-red color. As shown in Table 3, SVM classifier combined with Lab color space performed significantly better than other combinations that all the p -values are 0. This leads to conclude that SVM classifier combined with Lab color space is the optimal combination to classify pixels of fluorescent orange-red color in a construction scene.

Table 4 shows an empirical investigation of the effect of a color space transformation on the performance of the color-based safety vest detection by comparing performances of models with RGB color space and non-RGB color spaces (including nRGB). F-measure was used as a performance measure and the p -value was used to assess statistical significance. As shown in the Table 4, more than half of combinations presented increased value of F-measure especially for models based on Lab and YCbCr color spaces. For classification results on the data set of fluorescent yellow-green color, the performances of models based on the

Table 4. The Comparison Results between Classification Performances of RGB Color Space and non-RGB Color Spaces on the Data Sets of: (a) The Fluorescent Yellow-green Color, (b) The Fluorescent Orange-red Color

(a)						
	ANN	C4.5	KNN	LR	NB	SVM
RGB	95.26	98.01	97.94	92.99	75.47	95.29
nRGB	93.03 (0.000)*	95.34 (0.223)	95.74 (0.000)*	91.68 (0.000)*	87.80 (0.000)*	92.38 (0.000)*
HSV	94.30 (0.000)*	98.08 (0.012)*	97.91 (0.471)	74.30 (0.000)*	81.10 (0.013)*	96.00 (0.000)*
Lab	95.72 (0.038)*	98.08 (0.085)	97.93 (0.390)	94.03 (0.000)*	91.44 (0.000)*	98.05 (0.000)*
YCbCr	95.48 (0.013)*	98.11 (0.054)	97.99 (0.487)	93.01 (0.000)*	85.55 (0.000)*	97.69 (0.000)*
(b)						
	ANN	C4.5	KNN	LR	NB	SVM
RGB	94.95	98.18	98.21	90.95	79.92	96.75
nRGB	91.96 (0.901)	94.96 (0.000)*	94.64 (0.242)	87.54 (0.000)*	86.61 (0.000)*	88.06 (0.000)*
HSV	94.81 (0.000)*	98.27 (0.037)*	98.05 (0.188)	91.57 (0.000)*	82.75 (0.000)*	96.07 (0.000)*
Lab	95.02 (0.116)	98.17 (0.410)	98.20 (0.255)	90.34 (0.236)	91.58 (0.194)	98.49 (0.000)*
YCbCr	94.33 (0.000)*	98.20 (0.000)*	98.16 (0.041)*	90.95 (0.235)	86.01 (0.000)*	98.31 (0.000)*

Note. The F-measure values in bold indicates increased performance compared to the RGB color space. The numbers in parentheses are p -value. An asterisk indicates statistical significance (corresponding $p < 0.05$).

SVM presented increased values with significance by color space transformation except for nRGB color space. However, as presented on the Table 3, increased performances of models based on C4.5 was not statistically significant except for the HSV color space. For the models based on the KNN, the results indicated that color space transformation cannot help to improve a classification performance. On the data set of fluorescent orange-red color, the differences of models based on C4.5 and SVM had reached statistical significance. These results from

Table 5. The Comparison Results between Color Spaces including Luminance Components and excluding Luminance Components on the Data Sets of: (a) the Fluorescent Yellow-green Color, (b) the Fluorescent Orange-red Color

	ANN	C4.5	KNN	LR	NB	SVM
HSV	94.30	98.08	97.91	74.30	81.10	96.00
HS	91.76 (0.000)*	95.82 (0.000)*	95.78 (0.000)*	67.18 (0.577)	79.46 (0.000)*	93.69 (0.000)*
Lab	95.72	98.08	97.93	94.03	91.44	98.05
ab	94.18 (0.000)*	97.09 (0.000)*	97.40 (0.653)*	94.18 (0.061)	92.70 (0.003)*	95.82 (0.000)*
YCbCr	95.48	98.11	97.99	93.01	95.55	97.69
CbCr	94.69 (0.000)*	95.92 (0.000)*	95.92 (0.000)*	93.30 (0.461)	87.73 (0.000)*	95.72 (0.000)*

	ANN	C4.5	KNN	LR	NB	SVM
HSV	94.81	98.27	98.05	91.57	82.75	96.07
HS	92.61 (0.000)*	95.30 (0.000)*	94.58 (0.010)*	86.94 (0.000)*	83.73 (0.000)*	91.74 (0.059)
Lab	95.02	98.17	98.20	90.34	91.58	98.49
ab	92.77 (0.000)*	95.79 (0.080)	96.17 (0.045)*	89.08 (0.000)*	86.63 (0.000)*	95.54 (0.002)*
YCbCr	94.33	98.20	98.16	90.95	86.01	98.31
CbCr	93.50 (0.000)*	95.82 (0.036)*	95.65 (0.908)	90.23 (0.000)*	86.01 (0.000)*	95.71 (0.836)

Note. The values in bold indicates increased performance compared to the color spaces with luminance components. The numbers in parentheses are p -value. An asterisk indicates statistical significance (corresponding $p < 0.05$).

McNemar's test is corresponding to the results in Table 3. The results on the both data sets indicated that nRGB color space significantly degraded performance except for combination with NB. For classifying pixels of safety vests, these results strongly indicate that nRGB color space is not an optimal choice of a color space transformation.

An investigation of the role of the luminance components (V component of HSV, L component of Lab color space, and Y component of YCbCr) was also conducted in this study. As shown in Table 5, only the NB model with ab and HS color spaces improved classification performance. Except for these two models, the models based on color spaces with luminance components dominantly indicated better performance on both data sets. This leads to conclude that the classification performance relies on the role of luminance components. Moreover, this implies that the luminance components provide more information for classifying each color pixel of safety vests.

5. Conclusions

This study evaluated color-based safety vest detection in construction-site images obtained by means of the CMOS sensor. To find the best-performing color model and to investigate the effectiveness of color space transformation and the role of luminance components, a comparative analysis was conducted

on the performance of each model based on five color spaces (RGB, nRGB, HSV, Lab, YCbCr) combined with six classifiers (ANN, C4.5, KNN, LR, NB, and SVM). The classification performance was compared by three measures (precision, recall, and F-measure). McNemar's test was then applied to assess statistical significance among each performance of combinations.

The results presented that C4.5, KNN, and SVM classifiers achieved comparable performance on both data sets of fluorescent yellow-green and fluorescent orange-red colors. RGB color space transformation into non-RGB color spaces enhanced classification performance. On the contrary, normalizing RGB color space degraded the performance. According to the results of comparison between including and excluding luminance components in classification process, luminance components give more information in discriminating colors of safety vests and other backgrounds, thereby lead to better classification performance.

The results of this study, thus far, are encouraging and suggest that it would be more appropriate to use colors of safety vest as critical features in proceeding researches of detecting, localizing, tracking, and monitoring construction workers. When applying a color-based detection method, selecting the appropriate color space and classifier according to its application is a critical issue. However, in previous studies, color spaces and classifiers were determined without empirical analysis. The implications based on comparative analysis of this study provide guidance for selecting the color space and classifier in detecting safety vest on construction-site images and videos. Although the results of this study have shown a reliable classification performance, detection errors may occur since safety vest is detected using only color values. Future work will include more extensive and challenging data sets (i.e., colors of various objects in construction-site.) Moreover, use of other types of classifiers, such as a convolutional neural network, a recurrent neural network, and various types of ensemble classifier will be investigated. In addition, due to per-pixel model's high computational cost, a prior method is needed to reduce areas to be performed.

Acknowledgements

This research was supported by a grant (17SCIP-B079689-04) from Smart Civil Infrastructure Research Program funded by Ministry of Land, Infrastructure and Transport (MOLIT) of Korea government and Korea Agency for Infrastructure Technology Advancement (KAIA).

References

- Ali, S. and Smith, K. A. (2004). "On learning algorithm selection for classification." *Applied Soft Computing*, Vol. 6, No. 2, pp. 119-138, DOI: 10.1016/j.asoc.2004.12.002.
- ANSI/International Safety Equipment Association (ISEA) (2015). *American national standard for high-visibility safety apparel and headwear*, ANSI/ISEA 107-2015, Washington, DC., United States.
- Arlot, S. and Celisse, A. (2010). "A survey of cross-valid SS054.ation

- procedures for model selection.” *Statistics Surveys*, Vol. 4, pp. 40-79, DOI: 10.1214/09-SS054.
- Chang, C. C. and Lin, C. J. (2011). “LIBSVM: A library for support vector machines.” *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 3, DOI: 10.1145/1961189.1961199.
- Chen, M. Y. (2011). “Bankruptcy prediction in firms with statistical and intelligent techniques and a comparison of evolutionary computation approaches.” *Computers and Mathematics with Applications*, Vol. 62, No. 12, pp. 4514-2524, DOI: 10.1016/j.camwa.2011.10.030.
- Cheng, H. D., Jiang, X. H., Sun, Y., and Wang, J. (2001). “Color image segmentation: Advances and prospects.” *Pattern Recognition*, Vol. 34, No. 12, pp. 2259-2281, DOI: 10.1016/S0031-3203(00)00149-7.
- Ciatto, S., Houssami, N., Bernardi, D., Caumo, F., Pellegrini, M., Brunelli, S., Tuttobene, P., Bricolo, P., Fantò, C., Valentini, M., Montemezzi, S., and Macaskill, P. (2013). “Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): A prospective comparison study.” *The Lancet Oncology*, Vol. 14, No. 7, pp. 583-589, DOI: 10.1016/S1470-2045(13)70134-7.
- Delen, D., Walker, G., and Kadam, A. (2005). “Predicting breast cancer survivability: A comparison of three data mining methods.” *Artificial Intelligence in Medicine*, Vol. 34, No. 2, pp. 113-127, DOI: 10.1016/j.artmed.2004.07.002.
- Derelioglu, G. and Grgen, F. (2011). “Knowledge discovery using neural approach for SME’s credit risk analysis problem in Turkey.” *Expert Systems with Applications*, Vol. 38, No. 8, pp. 9313-9318, DOI: 10.1016/j.eswa.2011.01.012.
- Dreiseitl, S., Ohno-Machado, L., Kittler, H., Vinterbo, S., Billhardt, H., and Binder, M. (2001). “A comparison of machine learning methods for the diagnosis of pigmented skin lesions.” *Journal of Biomedical Informatics*, Vol. 34, No. 1, pp. 28-36, DOI: 10.1006/jbin.2001.1004.
- Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern classification 2nd Edition*, John Wiley and Sons.
- El Gayar, N., Schwenker, F., and Palm, G. (2006). “A study of the robustness of KNN classifiers trained using soft labels.” *Proc. Int. Conf. of the IAPR Workshop on Artificial Neural Networks in Pattern Recognition*, Springer, Ulm, Germany, pp. 67-80.
- Fritsch, J., Kuhn, T., and Geiger, A. (2013). “A new performance measure and evaluation benchmark for road detection algorithms.” *Proc. 16th Int. IEEE Conf. on Intelligent Transportation Systems*, IEEE, The Hague, Netherlands, pp. 1693-1700.
- Gong, J. and Caldas, C. H. (2011). “An object recognition, tracking, and contextual reasoning-based video interpretation method for rapid productivity analysis of construction operations.” *Automation in Construction*, Vol. 20, No. 8, pp. 1211-1226, DOI: 10.1016/j.autcon.2011.05.005.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*, Springer New York, USA.
- Holmes, C. C. and Adams, N. M. (2002). “A probabilistic nearest neighbour method for statistical pattern recognition.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 64, No. 2, pp. 295-306, DOI: 10.1111/1467-9868.00338.
- Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, John Wiley & Sons, NJ, USA.
- Jiang, B., Pan, Z., and Qiu, Y. (2017). “Study on the key technologies of a high-speed CMOS camera.” *Optik-International Journal for Light and Electron Optics*, Vol. 129, pp. 100-107, DOI: 10.1016/j.jileo.2016.10.056.
- Kakumanu, P., Makrogiannis, S., and Bourbakis, N. (2007). “A survey of skin-color modeling and detection methods.” *Pattern Recognition*, Vol. 40, No. 3, pp. 1106-1122, DOI: 10.1016/j.patcog.2006.06.010.
- Khan, R., Hanbury, A., Stttinger, J., and Bais, A. (2012). “Color based skin classification.” *Pattern Recognition Letters*, Vol. 33, No. 2, pp. 157-163, DOI: 10.1016/j.patrec.2011.09.032.
- Kim, D., Kim, D. H., Chang, S., Lee, J. J., and Lee, D. H. (2011). “Stability number prediction for breakwater armor blocks using Support Vector Regression.” *KSCE Journal of Civil Engineering*, Vol. 15, No. 2, pp. 225-230, DOI: 10.1007/s12205-011-1031-1.
- Liu, G. H. and Yang, J. Y. (2013). “Content-based image retrieval using color difference histogram.” *Pattern Recognition*, Vol. 46 No. 1, pp. 188-198, DOI: 10.1016/j.patcog.2012.06.001.
- MacInnes, J. J., Dickerson, K. C., Chen, N. K., and Adcock, R. A. (2016). “Cognitive neurostimulation: learning to volitionally sustain ventral tegmental area activation.” *Neuron*, Vol. 89, No. 6, pp. 1331-1342, DOI: 10.1016/j.neuron.2016.02.002.
- McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A., Andersen, G. L., Knight, R., and Hugenholtz, P. (2012). “An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea.” *The ISME Journal*, Vol. 6, No. 3, pp. 610-618, DOI: 10.1038/ismej.2011.139.
- 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*, Vol. 32, pp. 24-37, DOI: 10.1016/j.autcon.2012.12.002.
- Mohamed, A. S., Weng, Y., Ipson, S. S., and Jiang, J. (2007). “Face detection based on skin color in image by neural networks.” *Proc. Int. Conf. of Intelligent and Advanced Systems*, IEEE, Kuala Lumpur, Malaysia, pp. 779-783.
- Nallaperumal, K., Ravi, S., Babu, C. N. K., Selvakumar, R. K., Fred, A. L., Christopher, S., and Vinsley, S. S. (2007). “Skin detection using color pixel classification with application face detection: A comparative study.” *Proc. Int. Conf. on Computational Intelligence and Multimedia Applications*, IEEE, Sivakasi, Tamil Nadu, India, pp. 436-441.
- Park, M. W. and Brilakis, I. (2012). “Construction worker detection in video frames for initializing vision trackers.” *Automation in Construction*, Vol. 28, pp. 15-25, DOI: 10.1016/j.autcon.2012.06.001.
- Petri, M., Orbai, A. M., Alarcn, G. S., Gordon, C., Merrill, J. T., Fortin, P. R., Bruce, I. N., Isenberg, D., Wallace, D. J., Nived, O., Sturfelt, G., Ramsey-Goldman, R., Bae, S. C., Hanly, J. G., Snchez-Guerrero, J., Clarke, A., Aranow, C., Manzi, S., Urowitz, M., Gladman, D., Kalunian, K., Costner, M., Werth, V. P., Zoma, A., Bernatsky, S., Ruiz-Irastorza, G., Khamashta, M. A., Jacobsen, S., Buyon, J. P., Maddison, P., Dooley, M. A., van Vollenhoven, R. F., Ginzler, E., Stoll, T., Peschken, C., Jorizzo, J. L., Callen, J. P., Lim, S. S., Fessler, B. J., Inanc, M., Kamen, D. L., Rahman, A., Steinsson, K., Franks, A. G. Jr., Sigler, L., Hameed, S., Fang, H., Pham, N., Brey, R., Weisman, M. H., McGwin, G. Jr., and Magder, L. S. (2012). “Derivation and validation of the Systemic Lupus International Collaborating Clinics classification criteria for systemic lupus erythematosus.” *Arthritis and Rheumatism*, Vol. 64, No. 8, pp. 2677-2686, DOI: 10.1002/art.34473.
- Plataniotis, K. and Venetsanopoulos, A. N. (2013). *Color image processing and applications*, Springer Science & Business Media, Germany.
- Power, D. M. (2011). “Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation.” *Journal of Machine Learning Technologies*, Vol. 2, pp. 37-63.
- Quinlan, J. R. (2014). *C4.5: Programs for machine learning*, Elsevier, NY, USA.
- Radtke, J. P., Kuru, T. H., Boxler, S., Alt, C. D., Popeneciu, I. V.,

- Huettenbrink, C., Klein, T., Steinemann, S., Bergstraesser, C., Roethke, M., Roth, W., Schlemmer, H., Hohenfellner, M., and Hadaschik, B. A. (2015). "Comparative analysis of transperineal template saturation prostate biopsy versus magnetic resonance imaging targeted biopsy with magnetic resonance imaging-ultrasound fusion guidance." *The Journal of Urology*, Vol. 193, No. 1, pp. 87-94, DOI: 10.1016/j.juro.2014.07.098.
- Rice, W. R. (1989). "Analyzing tables of statistical tests." *Evolution*, Vol. 43, No. 1, pp. 223-225, DOI: 10.1111/j.1558-5646.1989.tb04220.x.
- Rish, I. (2001). "An empirical study of the Naïve Bayes classifier." *Proc. Int. Conf. of the IJCAI 2001 workshop on empirical methods in artificial intelligence*, IBM, pp. 41-46.
- Satio, T. and Rehmsmeier, M. (2015). "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets." *PloS one*, Vol. 10, No. 3, DOI: 10.1371/journal.pone.0118432.
- Seo, J. S., Han, S. U., Lee, S. H., and Kim, H. K. (2015). "Computer vision techniques for construction safety and health monitoring." *Advanced Engineering Informatics*, Vol. 29, No. 2, pp. 239-251, DOI: 10.1016/j.aei.2015.02.001.
- Shaik, K. B., Ganesan, P., Kalist, V., Sathish, B. S., and Jenitha, J. M. M. (2015). "Comparative study of skin color detection and segmentation in HSV and YCbCr color space." *Procedia Computer Science*, Vol. 57, pp. 41-48, DOI: 10.1016/j.procs.2015.07.362.
- Son, H., Hwang, N., Kim, C., and Kim, C. (2014). "Rapid and automated determination of rusted surface areas of a steel bridge for robotic maintenance systems." *Automation in Construction*, Vol. 42, pp. 13-24, DOI: 10.1016/j.autcon.2014.02.016.
- Son, H., Kim, C., and Kim, C. (2012). "Automated color model-based concrete detection in construction site images by using machine learning algorithms." *Journal of Computing in Civil Engineering*, Vol. 26, No. 3, pp. 421-433, DOI: 10.1061/(ASCE)CP.1943-5487.0000141.
- Son, H., Kim, C., Hwang, N., Kim, C., and Kang, Y. (2014). "Classification of major construction materials in construction environments using ensemble classifiers." *Advanced Engineering Informatics*, Vol. 28, No. 1, pp. 1-10, 2014, DOI: 10.1016/j.aei.2013.10.001.
- Taha, A. A. and Haunbury, A. (2015). "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool." *BMC Medical Imaging*, Vol. 15, No. 1, DOI: 10.1186/s12880-015-0068-x.
- Tarabalka, Y., Fauvel, M., Chanussot, J., and Benediktsson, J. A. (2010). "SVM-and MRF-based method for accurate classification of hyperspectral images." *IEEE Geoscience and Remote Sensing Letters*, Vol. 7, No. 4, pp. 736-740, DOI: 10.1109/LGRS.2010.2047711.
- Wei, Z., Wang, W., Bradfield, J., Li, J., Cardinale, C., Frackelton, E., Kim, C., Mentch, F., van Steen, K., Visscher, P. M., Baldassano, R. N., and Hakonarson, H., (2013). "Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease." *The American Journal of Human Genetics*, Vol. 92, No. 6, pp. 1008-1012.
- Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*, Morgan Kaufmann, Ma, USA.
- Zou, J. and Kim, H. (2007). "Using hue, saturation, and value color space for hydraulic excavator idle time analysis." *Journal of Computing in Civil Engineering*, Vol. 21, No. 4, pp. 238-246, DOI: 10.1061/(ASCE)0887-3801(2007)21:4(238).
- Zou, J., Kim, B., Kim, H., and Al-Hussein, M. (2012). "An automated system for the creation of an urban infrastructure 3D model using image processing techniques." *KSCE Journal of Civil Engineering*, Vol. 16, No. 1, pp. 9-17, DOI: 10.1007/s12205-012-1272-7.