



## Improving Detection Performance of Helmetless Motorcyclists Using the Combination of HOG, HOP, and LDB Descriptors

Sutikno<sup>1,2</sup>      Agus Harjoko<sup>2\*</sup>      Afiahayati<sup>2</sup>

<sup>1</sup>*Department of Computer Science, Diponegoro University, Indonesia*

<sup>2</sup>*Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia*

\* Corresponding author's Email: [aharjoko@ugm.ac.id](mailto:aharjoko@ugm.ac.id)

**Abstract:** A significant number of motorcyclists that do not wear helmets lose their lives during a traffic accident, one of the major causes of death globally. This led to the design and development of a system capable of detecting helmetless motorcyclists in real-time to reduce the number of deaths. Generally, this system consists of 3 subsystems, namely moving object segmentation, motorcycle classification, and helmetless head detection. The Histograms of Oriented Gradients (HOG) descriptor has been used in preliminary studies, which resulted in fast computation time and high accuracy. However, this descriptor was less effective when applied to images with varying lighting and was unable to distinguish local pattern features. Therefore, this study proposed a new descriptor algorithm, namely Histogram of Oriented Phase and Gradient- Local Difference Binary (HOPG-LDB), which combined the HOG, Histogram of Oriented Phase (HOP), and Local Difference Binary (LDB) descriptors. The HOP was used to enhance the inability of the HOG to be effectively used in detecting images with varying lighting, and the LDB was used to extract local pattern features. The results showed that the proposed method can improve the accuracy of motorcycle classification and helmetless head detection compared to HOG, HOP, LDB, HOG-HOP, HOG-LDB, and HOP-LDB descriptors. Furthermore, the motorcycle classification accuracies of the proposed method were 97.05%, 97.25%, and 99.35% for the JSC1, JSC2, and database1 datasets. Meanwhile, the helmetless head detection results of the proposed method were 71.21%, 66.63%, and 91.73 for the JSC1, JSC2, and database2 datasets.

**Keywords:** Motorcycle classification, Helmetless head detection, HOPG-LDB descriptor.

### 1. Introduction

A traffic accident is among the top 10 causes of death globally [1]. According to analysis, as many as 1.35 million people die yearly due to traffic accidents [1]. In Indonesia, riders of motorcycles and tricycles make up 74% of all traffic accident-related deaths [1]. The most common cause of death in this group is head injuries, usually due to their inability to wear helmets. Although most countries have mandated the use of helmets, there are still many motorcyclists that violate these rules. Therefore, a system capable of automatically detecting helmetless motorcyclists in real-time is needed.

Several studies have proposed several methods for the detection of a helmetless motorcyclist with accuracy and computation time as a measure of

success. The feature extraction process is one of the determining factors of this success. In general, there are 2 methods, namely hand-crafted features and Convolution Neural Network (CNN). The CNN produced better accuracy on large training data, but the computation time is longer [2]. Meanwhile, the hand-crafted features have faster computation time and relatively good accuracy [3].

Histograms of Oriented Gradients (HOG) descriptor is a hand-crafted feature that produces better performance than several descriptors. For instance, Silva et al. proposed a detection system made into 2 subsystems, namely motorcycle detection and helmetless head detection [4]. The subsystem of helmetless head detection compared Wavelet Transform (WT), HOG, Local Binary Pattern (LBP) descriptors, and a combination of all

descriptors. The result showed that the HOG descriptor has higher accuracy than others.

Dahiya et al. compared HOG, Scale-Invariant Feature Transform (SIFT), and LBP descriptors in the subsystem of motorcycle detection and helmetless head detection [5]. The result showed that the HOG descriptor also has higher accuracy than others.

Moreover, the HOG descriptor has been developed to tackle some issues. For instance, Ragb & Asari proposed a Histogram of Oriented Phase and Gradient (HOPG) descriptor that combined HOG and Histogram of Oriented Phase (HOP) to detect the presence of pedestrians on roads [6]. HOP was added to make up for the weakness of HOG, which was ineffective in detecting images of varied lighting and uneven background. The result showed that their proposed descriptor improved detection accuracy by 24.92% from the HOG. Hua Wang et al. combined HOG and Local Difference Binary (LDB) for face recognition [7]. The LDB descriptor was added to make up for the weakness of HOG, which was incapable of differentiating images with varying local patterns. The result showed that the HOG-LDB descriptor has higher accuracy than the HOG.

This study aims to develop a new method of hand-crafted features that can increase accuracy. The proposed method combines gradient, local phase, and pattern features. The subjects consist of a dataset of images with varied lighting, uneven background, and different local patterns. The gradient-based, local phase-based, and local pattern descriptors are HOG, HOP, and LDB, respectively. The combination algorithm denoted Histogram of Oriented Phase and Gradient - Local Difference Binary (HOPG-LDB) descriptor for motorcycle classification and helmetless head detection.

The rest of the paper is organized as follows: Section 2 reviews related work. Section 3 describes the proposed method. Section 4 presents experiments and results, and Section 5 elaborates the conclusion of the study.

## 2. Related work

Generally, this study is divided into 2 subsystems, namely motorcycle detection and helmetless head detection. The following sections explain the types of feature extractions used for both subsystems.

### 2.1 Motorcycle detection

Motorcycle detection has involved 3 processes, namely vehicle segmentation, feature extraction, and classification. Feature extraction for motorcycle

detection includes 4 types, namely shape, texture, combination (shape, color, and texture), and CNN.

Shape feature extraction is the most common method used to differentiate motorcycles and other vehicles. Leelasantitham and Wongseree used 3 shape features, namely length, width, and their ratios to classify vehicles into 5 groups [8]. The results obtained from using the decision tree classifier showed high accuracy, however, the features were not capable of differentiating bicycles, motorcycles, and tricycles. Fazli et al. used the features of length, width, area, diameter, and the ratio of distance to determine the object's center of mass and its main axis length [9]. The classification process used a multilayer perceptron (MLP) to classify the vehicles into 3 categories, namely heavy and light duties, and motorcycles.

Marayatr et al. used the area feature to classify motorcycles and other vehicles [10]. This feature was classified using MLP. Meanwhile, Dupuis et al. proposed a method focusing on calculating the number of motorcycles on the road in real-time [11]. This analysis used area, height, and width features to classify motorcycles and non-motorcycles.

Dahiya et al. compared some descriptors, namely HOG, SIFT, and LBP in classifying motorcycles and non-motorcycles. The results showed that the HOG descriptor has the best accuracy [5]. Ashvini et al. compared HOG, Speeded Up Robust Features (SURF), SIFT, and LBP in motorcycles detection [12]. This study also used data of images taken from in front, beside, and behind motorcycles. The results showed that the HOG descriptor has higher accuracy.

A study by Chen et al. proposed a system that classified vehicles into 4 categories, namely cars, vans, buses, and motorcycles [13]. The system used Intensity Pyramid-based HOG (IPHOG) descriptor and support vector machine (SVM) classifier. The results showed that the conditions of weather and light changing have lower accuracy than the normal condition.

Silva et al. used the LBP descriptor and SVM classifier to detect motorcycles [14]. This descriptor was compared with SURF, HOG, and Haar Wavelet. The results showed that it has better accuracy than the others. The descriptor does not require setting initial parameters, such as camera angle and height, and vehicle size. Silva et al. proposed a WT descriptor that was compared with LBP, HOG, and SURF [4]. The results showed that it has better accuracy than the others.

Waranusast et al. proposed a combination of shape and color features comprising of area, the ratio of width and height, and color deviation standard [15]. These characteristics served as input for the k-nearest

neighbors (KNN) classifier to determine the motorcycles and non-motorcycles. The proposed technique was able to calculate the number of passengers on a motorcycle. The results showed errors in classification because the data were taken from afar, overlapping vehicles, and the passenger sitting too close to the rider. Shine and Jiji concatenated HOG and LBP with sequential minimum optimization (SMO) for training the SVM classifier. The results show that the combination of these descriptors produced better accuracy than the HOG and LBP descriptors [3].

CNN has also been used for motorcycle detection, in which the characteristics were observed using convolution operation on the input images. Huynh et al. proposed a detection method using CNN, focusing on traffic jam situations [2]. Vishnu et al. also used CNN to deal with various lighting and poor video quality [16]. The test results show that the accuracy generated by CNN is higher than hand-crafted features, but the computation time is much longer.

## 2.2 Helmetless head detection

Helmetless head detection has involved 3 stages, namely ROI (region of interest) determination, feature extraction, and classification. The ROI determination aims to ascertain the area around a rider's head. The heads of the rider and passenger are above the motorcycle image, therefore, the research focused on the upper part of the object. Once the head area is known, the next steps are feature extraction and classification. Feature extraction for helmetless head detection includes 3 types, namely shape, combination (shape, texture, and color), and CNN.

Marayatr and Kumhon used the circular hough transform (CHT) descriptor for helmetless head detection. The result showed that it was found erroneous for the detection of 2 or more passengers [10]. Silva et al. proposed the HOG descriptor, and the dataset for the study was taken in a static environment [4]. The assessment was performed by comparing HOG, WT, LBP, WT+LBP, WT+HOG, LBP+HOG, and WT+HOG+LBP descriptors. The results showed that the HOG descriptor has the best accuracy.

Dahiya et al. compared HOG, SIFT, and LBP descriptors [5]. The results showed that HOG has the best accuracy. The data used were obtained on a relatively quiet road, therefore, this method has not been tested in more complex environments.

Some studies have combined shape, texture, and color features to improve accuracy. Waranusast et al. used features of arc circularity, average intensity, and

hue [15]. This study was performed on 3 recording conditions, such as near, far, and medium. It was found that the greatest errors were from the data recording obtained from afar.

Ashvini et al. used the features of arc circularity, average intensity and hue, and Center Symmetric-Local Binary Pattern (CS-LBP) of the head [12]. These features served as input to the KNN classifier for the classification of heads with and without helmets. The method focused on issues with data recording taken from different angles in the front, beside, and at the back. The head images taken were cropped manually. Talaulikar et al. also used arc circularity, average intensity and color, and HOG [17].

Silva et al. used geometric, shape, and texture characteristics [14]. The study used a combination of CHT, LBP, and HOG descriptors. CHT was used to determine the geometric shape of an image. The results obtained from LBP and HOG feature extraction were used as input for the classification process. However, this method has a weakness of not being able to detect images of low resolution.

Mistry et al. used CNN with the YOLOv2 model to detect riders without helmets [18]. Meanwhile, Vishnu et al. used the AlexNet model on both light and heavy traffic [16]. Raj KC et al. also used the AlexNet model and erroneous detections were made for riders putting on hats [19]. Forero used the iter\_45, Inception-V3 network, and full ImageNet network models [20]. Mayya and Nayak proposed a technique of reducing the computing time, using Faster Regions with Convolution Neural Network (R-CNN) [21]. However, erroneous detections were still made for bicycle riders. Shine and Jiji proposed hand-crafted features (a combination of HOG, LBP, and Haralick) and Custom CNN [3]. The test results show that Custom CNN is superior in terms of accuracy and hand-crafted features are superior in terms of prediction time with relatively good accuracy.

## 3. The proposed method

The following section explains the proposed method to detect helmetless motorcyclists. This method is divided into 3 subsystems, namely moving object segmentation, motorcycle classification, and helmetless head detection. Fig. 1 shows the stages of the proposed method.

### 3.1 Moving object segmentation

The segmentation of moving objects is used to separate them from the background. The output is the images comprising of motorcycles, cars, bikes,

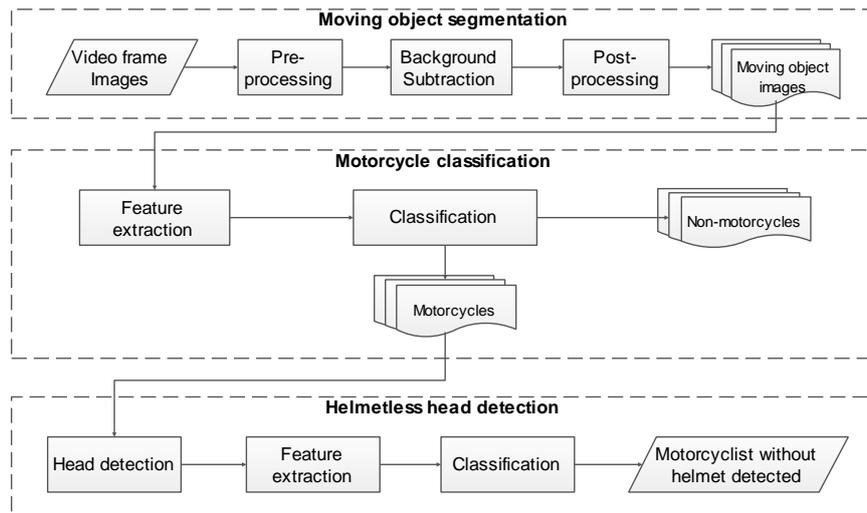


Figure. 1 The stages of the proposed method

people, etc. This subsystem is further divided into pre-processing, background subtraction, and post-processing. Pre-processing prepared each frame for the process of background subtraction. It comprises ROI setting, grayscale image conversion, and contrast enhancement. The setting of ROI aimed to reduce the frame size to allow quicker computation. Users manually set ROI for road areas where a large number of vehicles pass. Fig. 2 shows an example of a video frame and road area set by users. The yellow area was determined by choosing 4 points of road barrier. After setting the area, the ROI frame and virtual line (red) were obtained. The endpoints of the virtual line were set in the middle of both road barriers. Fig. 3 shows a sample of the ROI frame and virtual line. The following process is grayscale image conversion and contrast enhancement using histogram equalization.

The background subtraction process employed the Gaussian Mixture Model capable of adapting to varied lighting conditions [22]. The output was blobs of moving objects, along with noise, which was corrected using post-processing. Post-processing was divided into filtering and morphological operation (opening and dilation). The filtering process aimed to determine moving objects, especially motorcycles. It was carried out by limiting the height and width of blobs. Therefore, besides from that, the blobs with the diagonal of the bounding box crossing the virtual line would be used for further processing.

### 3.2 Motorcycle classification

The subsystem of motorcycle classification is used to classify moving objects into motorcycles and non-motorcycles. It is divided into feature extraction and classification.

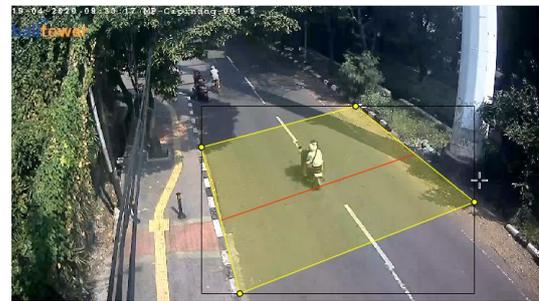


Figure. 2 Sample of frame

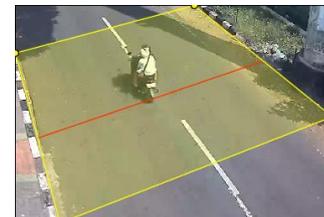


Figure. 3 Sample of ROI frame

#### 3.2.1. Feature extraction

Feature extraction is used to extract the attributes of the moving object image. This process was carried out using the HOPG-LDB descriptor that combined HOG, HOP, and LDB descriptors.

##### 3.2.1.1. HOG descriptor

HOG is a window-based descriptor that detects points of interest. Each image has characteristics shown by gradient distribution, which is gained from dividing images into cells. Furthermore, each cell is made of a gradient histogram. A combination of these histograms is used to represent an object [23].

The initial stage to obtaining the HOG feature was by calculating horizontal ( $I_x$ ) and vertical gradients ( $I_y$ ) in sequence, as shown in Eqs. (1) and (2).

$$I_x = I \times D_x \quad (1)$$

$$I_y = I \times D_y \quad (2)$$

where  $I$  denotes a gray level image,  $D_x$  is mask  $[-1 \ 0 \ 1]$ , and  $D_y$  is mask  $[-1 \ 0 \ 1]^T$ .  $I_x$  and  $I_y$  are used to calculate gradient magnitude ( $|G_r|$ ) and orientation ( $\theta$ ) in sequence, as shown in Eqs. (3) and (4).

$$|G_r| = \sqrt{I_x^2 + I_y^2} \quad (3)$$

$$\theta = \tan^{-1} \left( \frac{I_y}{I_x} \right) \quad (4)$$

HOG is applied in the image that is divided into some blocks and further into cells. Gradient orientation histogram is calculated for each cell with 9-orientations binning.

### 3.2.1.2. HOP descriptor

HOP is a descriptor based on phase congruency (PC) values and orientation angles of each image pixel. Kovasi developed the PC technique to detect tips and edges of digital images [24]. It is calculated by extracting local frequency and phase information through the combination of image input using a quadrature filter. One of the most efficient filters is the Log-Gabor with its transfer function outlined in Eq. (5) [25].

$$G(\omega) = \exp \left( \frac{-(\log(\omega/\omega_0))^2}{2(\log(k/\omega_0))^2} \right) \quad (5)$$

where  $\omega_0$  is the middle frequency of the filter and  $k/\omega_0$  stays constant for varied  $\omega_0$ . The transfer function of 2D Log-Gabor in the angular direction is given in Eq. (6) [25].

$$G(\theta) = \exp \left( \frac{-(\theta-\theta_0)^2}{2\sigma_\theta^2} \right) \quad (6)$$

where  $\theta_0$  denotes the middle orientation of the filter, and  $\sigma_\theta$  is the standard deviation of the Gaussian function in the angular direction.

The response vector at scale  $n$  and orientation  $o$  is calculated by convolution of each quadrature pair with the input signal  $I(x,y)$ , as shown in Eq. (7) [6].

$$\begin{aligned} [e_{no}(x,y), o_{no}(x,y)] \\ = [I(x,y) * M_{no}^e, I(x,y) * M_{no}^o] \end{aligned} \quad (7)$$

where  $M_{no}^e$  and  $M_{no}^o$  denote the even-symmetric and odd-symmetric filter at scale  $n$  and orientation  $o$ .

The amplitude of the response  $A_{no}$  is calculated using Eq. (8).

$$A_{no} = \sqrt{(e_{no}^2(x,y)) + o_{no}^2(x,y)} \quad (8)$$

$F(x,y)$  is the input signal  $I(x,y)$  filtered from a DC component for 2D Signal and  $F_H(x,y)$  is  $90^\circ$  phase shift of  $F(x,y)$  (Hilbert Transform).  $F(x,y)$  and  $F_H(x,y)$  can be defined as in Eqs. (9) and (10).

$$F(x,y) = \sum_o \sum_n e_{no}(x,y) \quad (9)$$

$$F_H(x,y) = \sum_o \sum_n o_{no}(x,y) \quad (10)$$

Therefore, phase congruency  $PC(x,y)$  and orientation angle  $\varphi(x,y)$  of the 2D signal can be calculated using Eqs. (11) and (12).

$$PC(x,y) = \frac{\sum_o \sqrt{(\sum_n e_{no}(x,y))^2 + (\sum_n o_{no}(x,y))^2}}{\varepsilon + \sum_o \sum_n A_{no}(x,y)} \quad (11)$$

$$\varphi(x,y) = \tan^{-1} \left( \frac{F_H(x,y)}{F(x,y)} \right) \quad (12)$$

where  $\varepsilon$  is a very small positive real number, used to prevent division of zero. This descriptor is applied by dividing the image into some blocks and cells. The phase orientation histogram is calculated for each cell using 9-orientation binning.

### 3.2.1.3. LDB descriptor

LDB was first introduced by Yang and Cheng [26]. It was inspired by the self-similarity descriptor, which divides images into grids of the same size using the relationships between the central and remaining grids.

The first step in calculating LDB features is dividing images into blocks, with each further split into cells. This is followed by calculating the integral image of each block. The result of the integral image of size  $N \times N$  is  $(N+1) \times (N+1)$ . Fig. 4 shows the original image size, while Fig. 5 indicates its integral result.

The next step is calculating the sum of pixel intensity ( $I_s$ ), an average of pixel intensity ( $\bar{I}$ ), horizontal gradient ( $I_x$ ), and vertical gradient ( $I_y$ ) of each cell, using Eqs. (13) to (16).  $I_1, I_2, I_3, I_4, I_5, I_6, I_7$ , and  $I_8$  are values of the integral image of each cell, as shown in Fig. 6.

$$I_s = i_1 + i_4 - i_3 - i_2 \quad (13)$$

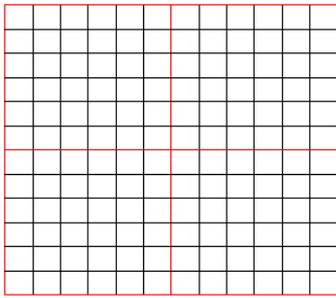


Figure. 4 Original image

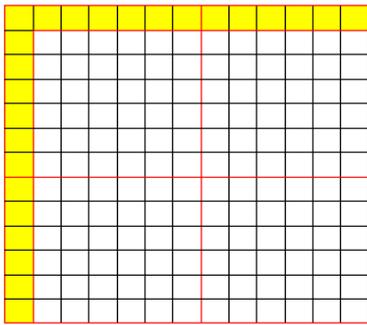


Figure. 5 Result of integral image



Figure. 6 The integral image of a cell

$$\bar{I} = \frac{1}{n^2} \times I_s \tag{14}$$

$$I_x = I_s - 2 \times (i_1 + i_7 - i_5 - i_3) \tag{15}$$

$$I_y = I_s - 2 \times (i_1 + i_6 - i_2 - i_8) \tag{16}$$

The final step is determining 3 binary string bits to compare every pair of cells on each block using Eq. (17).

$$F(f(i), f(j)) = \begin{cases} 1, & f(i) - f(j) > 0 \text{ and } i \neq j \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

where  $f(i)$  and  $f(j)$  are functions of information extraction from each cell. The function has the value of  $\bar{I}$ ,  $dx$ , and  $dy$ . LDB features are determined by combining binary string bits of all blocks.

### 3.2.1.4. Combination of HOG, HOP, and LDB descriptors

A combination of HOG, HOP, and LDB forms the HOPG-LDB descriptor. This descriptor is the proposed method used to detect motorcyclists

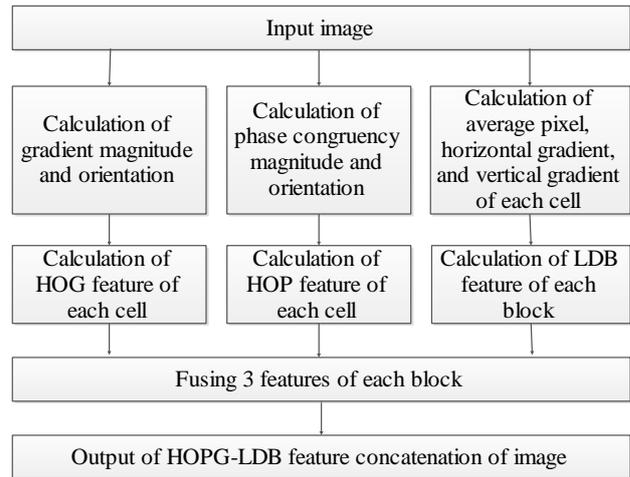


Figure. 7 Constructing HOPG-LDB descriptor

without a helmet. HOG and HOP are used to capture features of image shape, while LDB captures image local patterns [6-7]. Fig. 7 shows the combining process of these descriptors.

HOPG-LDB descriptor is implemented by dividing input images into blocks of the same size without any overlap. Each block is made up of some cells. Values of gradient magnitude and orientation and PC magnitude and orientation are calculated for each pixel. These values are further used to make HOG and HOP on each cell and block. A sample of HOG and HOP construction on the first block is shown in Eqs. (18) and (19), respectively, where  $t$  denotes the number of cells in each block.

$$HOG_{b1} = [HOG_{c1} HOG_{c2} \dots HOG_{ct}] \tag{18}$$

$$HOP_{b1} = [HOP_{c1} HOP_{c2} \dots HOP_{ct}] \tag{19}$$

The HOG and HOP for other blocks are calculated using similar methods, while the LDB feature is determined for each block. Fig. 8 shows an illustration of the 3 features combination, where  $u$  is the number of blocks.

The image input of this descriptor is 96×48 pixels. This study tested varying sizes of block and cell before selecting the best. Block size variations used were 2×2 and 3×3. Meanwhile, cell size variations used for the 2×2 were 4×4, 6×6, 8×8, and 12×12 pixels. Conversely, cell size variations used for the 3×3 were 4×4 and 8×8 pixels.

### 3.2.2. Classification

The MLP classifier is used to categorize objects into motorcycles and non-motorcycles. The variations of the number of hidden layers were tested, and the training algorithm with the best result was selected. The number of hidden layer variations used

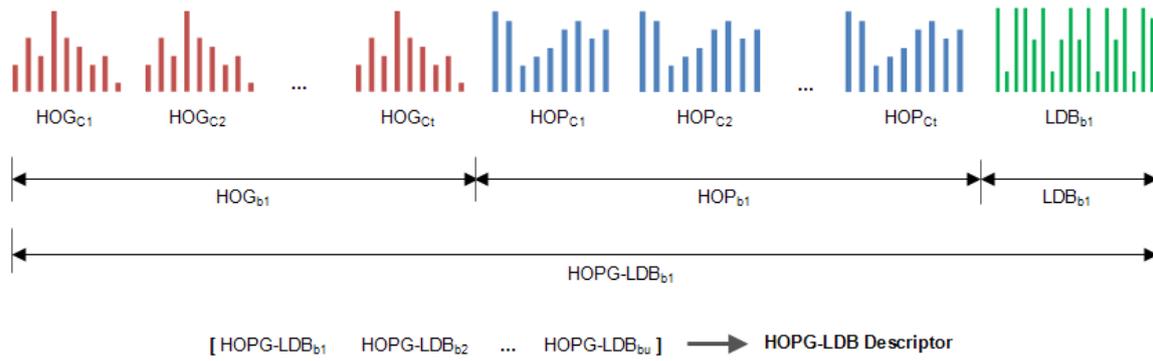


Figure. 8 An illustration of HOPG-LDB descriptor construction

was 1, 2, and 3. In addition, the variation of the number of neurons on hidden layers ( $n_H$ ) was determined by using a formula put forward by Chen et al., in Eq. (20) [27].

$$n_H = \sqrt{n_i + n_o} + l \quad (20)$$

where  $n_i$  and  $n_o$  denotes the number of neurons on the input and output layers, while  $l$  is an integer constant of 1 to 10. The variations of the MLP training algorithm consisting of backpropagation of Gradient descent with adaptive learning rate and scaled conjugate gradient, conjugate gradient backpropagation with Powell-Beale restarts, conjugate gradient backpropagation with Fletcher-Reeves update, conjugate gradient backpropagation with Polak-Ribière update, one step secant backpropagation, gradient descent with momentum and adaptive learning rate backpropagation, and gradient descent backpropagation. The parameter value for training used learning rate, the maximum number of an epoch, and limit for error are 0.05, 1000, and 0.001, respectively [28].

### 3.3 Helmetless head detection

The helmetless head detection subsystem is divided into head detection, feature extraction, and classification.

#### 3.3.1. Head detection

The input of the head detection process is a motorcycle image, while the output is the detected

head of the rider and passenger. The first step was to determine the ROI of the head by analyzing the minimum and maximum positions of the upper 1/3 of the blob image generated by the segmentation. These positions are used as the ROI head limit, converted to grayscale, and contrasted using contrast-limited adaptive histogram equalization (CLAHE). An example of a motorcycle image and the results of this process are shown in Fig. 9(a) and 9(b), respectively.

The next step was to create 2 images using thresholding and inverse thresholding to produce 2 images with opposite intensities, as shown in Fig. 9(c). This figure shows that there were still useless blobs that need to be filtered and corrected. The filtering is carried out by morphological operations (opening and filling holes), removing blobs on the sides and top, and filtering based on wide and high thresholds. The blob is fixed by cutting the bottom part with a certain threshold. The result of this process is shown in Fig. 9(d).

The next process was edge detection and the application of CHT, as shown in Fig. 9(e) and 9(f). This process was carried out using the Laplace of Gaussian (LoG) edge detection. The circulars on these two images were combined in the ROI head image, as shown in Fig. 9(g). Each bounding box on the circular object was classified to determine the head object. The classification process used the MLP classifier and the HOPG-LDB descriptor. The result of this process is shown in Fig. 9(h). The last step was to combine several detections supposing one object has more than one by using non-maximum suppression (NMS).

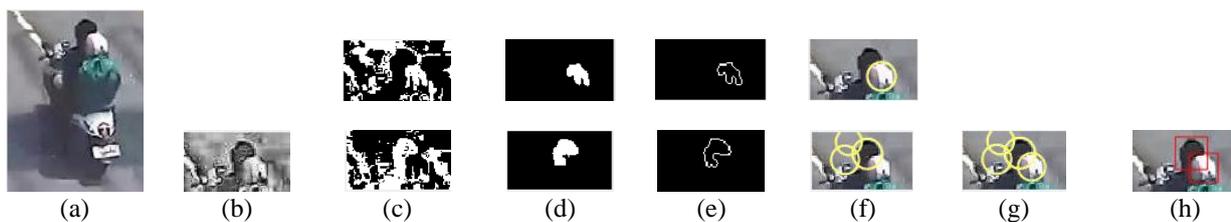


Figure. 9 Image sample of results from each step: (a) motorcycle input, (b) ROI head determination, (c) thresholding, (d) filter and fix blob, (e) edge detection, (f) CHT application, (g) combine CHT, and (h) classification

### 3.3.2. Feature extraction

The HOPG-LDB descriptor is used for feature extraction, and its input is an image of 48×48 pixels. Testing was performed by varying block and cell size. Block size variations used were 2×2 and 3×3, while 2×2, 4×4, 6×6, 8×8, and 12×12 pixels were used for a cell size of 2×2. Cell size variations used for 3×3 were 2×2, 4×4, and 8×8 pixels. The best result of the experiment is selected.

### 3.3.3. Classification

Classification is used to separate the heads with and without helmets. The classifier used was MLP, which was tested by varying the number of hidden layers, neurons, and the training algorithm. These variation values are the same as the various values used in the motorcycle classification subsystem.

### 3.4 Performance measurement

Parameters used to measure the performance of the motorcycle classification subsystem were accuracy (Acc), precision (Pre), and recall (Rec). Meanwhile, those used to measure the performance of the helmetless head detection subsystem were average precision (AP), precision, and recall. AP was the most common parameter used to measure detection accuracy [29]. Especially for the *database1* and *database2* datasets, the experiments used the K-Fold cross-validation method (K=10), and performance measurement used accuracy, precision, and recall to adjust the previous study [4]. Equations used to calculate accuracy, precision, and recall are Eqs. (21) and (23), respectively.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (21)$$

$$Precision = \frac{TP}{TP+FP} \quad (22)$$

$$Recall = \frac{TP}{TP+FN} \quad (23)$$

where TP denotes a true positive, TN represents a true negative, FP is a false positive, and FN is a false negative. TP is the right detection of the ground-truth bounding box. Correct detection is measured with intersection over union (IOU), using Eq. (24).

$$IOU = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \quad (24)$$

where  $B_p$  is the prediction bounding box, and  $B_{gt}$  is the ground-truth bounding box. Detection is deemed

correct if the value of IOU  $\geq t$ . The  $t$  value for this study is 0.5.

AP is an area below the precision-recall curve of between 0 and 1. The equation for AP is Eq. (25).

$$AP = \sum_n (R_{n+1} - R_n) P_i(R_{n+1}) \quad (25)$$

where  $P_i(R_{n+1})$  is calculated by Eq. (26).

$$P_i(R_{n+1}) = \max_{\tilde{R}: \tilde{R} \geq R_{n+1}} P(\tilde{R}) \quad (26)$$

where  $P(\tilde{R})$  is the precision measured during recall  $\tilde{R}$ .

## 4. Experiments and results

### 4.1 Dataset

This study used two private datasets and two public datasets. These private datasets are *JSC1* and *JSC2* taken from the rear and front of vehicles, respectively. The datasets consist of motorcycle and not-motorcycle images that are used as an input for the motorcycle classification subsystem. The images of the motorcycle are used as input to the helmetless head detection subsystem. Examples of *JSC1* and *JSC2* datasets are shown in Fig. 10 and Fig. 11, respectively.

The *JSC1* dataset was generated from CCTV (Closed Circuit Television) video recordings on Cipinang Baru Timur Street, East Jakarta, with a frame rate of 19.49 fps. Meanwhile, the *JSC2* dataset was generated from CCTV video recordings on Budi Raya Street, West Jakarta, with a frame rate of 20 fps. The CCTVs are owned by the Jakarta Smart City (JSC), Jakarta province government, Indonesia. Both videos have a resolution of 1280 x 720 pixels and a duration of 3 hours. Training data is taken from the first 2 hours and testing data from the rest. The previous study also used this data division technique to detect motorcyclists without helmets [5]. The number of the training and testing data used in *JSC1* and *JSC2* datasets are shown in Table 1 and Table 2, respectively.



Figure 10 Samples of *JSC1* dataset



Figure. 11 Samples of *JSC2* dataset

Table 1. The number of training data

Subsystem	Image	<i>JSC1</i> dataset	<i>JSC2</i> dataset
Motorcycle classification	Motorcycle	1602	4066
	Non-motorcycle	1602	4066
	Total	3204	8132
Helmetless head detection	Head with helmet	2052	1984
	Head without helmet	1694	1984
	Total	3746	3968

Table 2. The number of testing data

Subsystem	Image	<i>JSC1</i> dataset	<i>JSC2</i> dataset
Motorcycle classification	Motorcycle	531	1390
	Non-motorcycle	587	1852
	Total	1118	3242
Helmetless head detection	Motorcyclist with helmet	416	1091
	Motorcyclist without helmet	115	299
	Total	531	1390

This study also used public datasets, namely *database1* and *database2* [4]. *Database1* is used for the motorcycle classification subsystem and *database2* is used for the helmetless head detection subsystem. *Database1* contains 2576 non-motorcycle images and 669 motorcycle images. *Database2* contains the images of the head's motorcyclist generated by determining the ROI process. These images consist of 151 heads with helmets and 104 heads without a helmet. Examples of *database1* and *database2* datasets are shown in Fig. 12 and 13.



Figure. 12 Samples of *database1*



Figure. 13 Samples of *database2*

## 4.2 Results and discussion

This section explains the testing results of the proposed method for motorcycle classification and helmetless head detection. The proposed method is also compared to the previous studies.

### 4.2.1. Motorcycle classification

The comparison result of descriptors for motorcycle classification is shown in Table 3. It illustrates that the accuracies of the HOP are higher than the HOG and LDB for the *JSC1* dataset, and the accuracies of HOG are higher than the HOP and LDB for *JSC1* and *database1* datasets. Accuracies of HOG-LDB are higher than the single descriptor and the combination of the other two for *JSC1* and *database1* datasets. Furthermore, the combination of HOG, HOP, and LDB (HOPG-LDB) produces the best performance on all datasets. The accuracy, precision, and recall of the HOPG-LDB in the *JSC1* dataset were 97.05%, 96.98%, and 96.80%, respectively. The accuracy, precision, and recall of HOPG-LDB in the *JSC2* dataset were 97.25%, 97.38%, and 96.19%, respectively. Meanwhile, the accuracy, precision, and recall of HOPG-LDB in the *database1* dataset were 99.35%, 98.52%, and 98.36%, respectively. Therefore, it can be concluded that the combination of HOG, HOP, and LDB contains important information that contributes to improving motorcycle classification performance.

The comparison result of the proposed method and previous studies is shown in Table 4. Here, [5] used HOG+SVM combination, [4] used WT+RF combination, and [3] concatenated HOG and LBP descriptors with SMO classifier. The result showed that the proposed method has better accuracy than these previous studies for all datasets.

### 4.2.2. Helmetless head detection

Table 5 shows the comparison result of descriptors for helmetless head detection. This table shows that LDB produces higher AP than other single descriptors for *JSC1* and *JSC2* datasets, and HOG produces higher AP than other single descriptors for *database2* datasets. HOG-LDB produces higher AP than other single and double descriptors for *JSC1* and *JSC2* datasets. The combination of HOG, HOP, and LDB results in the best performance on all datasets. For the *JSC1* dataset, HOPG-LDB produces an AP of 71.21%, a precision of 58.92%, and a recall of 80.74%. In the *JSC2* dataset, it produces an AP of

Table 3. Comparison of descriptors for motorcycle classification

Descriptor	JSCI dataset			JSC2 dataset			database1 dataset		
	Acc (%)	Pre (%)	Rec (%)	Acc (%)	Pre (%)	Rec (%)	Acc (%)	Pre (%)	Rec (%)
HOG	93.83	95.47	91.34	96.85	97.63	94.96	98.71	98.02	96.71
HOP	94.45	95.01	93.22	90.90	91.57	86.76	98.18	96.70	95.52
LDB	91.32	90.79	90.96	90.90	90.65	87.84	96.68	95.12	91.47
HOG-HOP	96.87	97.69	95.67	97.10	98.43	94.75	98.89	98.21	97.31
HOG-LDB	96.96	95.93	97.74	97.07	97.93	95.18	98.98	98.05	97.91
HOP-LDB	95.80	96.01	95.10	96.27	97.67	93.53	98.58	97.10	95.56
<b>HOPG-LDB (Proposed method)</b>	<b>97.05</b>	<b>96.98</b>	<b>96.80</b>	<b>97.25</b>	<b>97.38</b>	<b>96.19</b>	<b>99.35</b>	<b>98.52</b>	<b>98.36</b>

Table 4. Comparison of the proposed method and previous studies for motorcycle classification

Method	JSCI dataset			JSC2 dataset			database1 dataset		
	Acc (%)	Pre (%)	Rec (%)	Acc (%)	Pre (%)	Rec (%)	Acc (%)	Pre (%)	Rec (%)
[5]	94.81	93.08	96.23	96.92	98.35	94.39	98.58	97.60	95.52
[4]	96.78	96.61	96.61	91.73	89.79	91.08	97.78	97.16	91.93
[3]	93.92	90.40	97.55	95.87	97.50	92.73	99.26	98.67	97.76
<b>Proposed method</b>	<b>97.05</b>	<b>96.98</b>	<b>96.80</b>	<b>97.25</b>	<b>97.38</b>	<b>96.19</b>	<b>99.35</b>	<b>98.52</b>	<b>98.36</b>

66.63%, a precision of 53.79%, and a recall of 82.46%. Meanwhile, in the *database2* dataset, it produces an AP of 91.73%, a precision of 92.00%, and a recall of 88.27%. Therefore, the combination of HOG, HOP, and LDB contains essential information that contributes to improving helmetless head detection performance for all datasets.

The results of the comparison of the proposed method and previous studies are shown in Table 6.

Here, [5] used HOG+SVM combination, [4] used HOG+MLP combination, and [3] used Custom CNN. From the data in Table 6, we can be seen that our method is superior for *JSCI* and *database2* datasets, while Custom CNN in [3] is superior for the *JSC2* dataset. The precision-recall curves of the descriptors are shown in Fig. 14. In addition, precision-recall curves of the proposed method and previous studies are shown in Fig. 15.

Table 5. Comparison of descriptors for helmetless head detection

Descriptor	JSCI dataset			JSC2 dataset			database2 dataset		
	AP (%)	Pre (%)	Rec (%)	AP (%)	Pre (%)	Rec (%)	Acc (%)	Pre (%)	Rec (%)
HOG	61.50	54.84	75.56	55.03	43.19	80.87	88.67	89.75	85.55
HOP	61.43	52.11	73.33	57.52	43.25	82.46	85.06	85.05	82.36
LDB	62.34	50.24	78.52	58.19	45.31	78.13	86.20	86.37	85.45
HOG-HOP	59.50	57.87	76.30	61.83	38.73	84.97	88.66	88.61	86.27
HOG-LDB	67.91	57.30	78.52	65.75	52.96	81.55	87.46	90.67	83.55
HOP-LDB	66.02	50.00	79.26	65.48	49.59	82.00	88.58	92.82	81.45
<b>HOPG-LDB (Proposed method)</b>	<b>71.21</b>	<b>58.92</b>	<b>80.74</b>	<b>66.63</b>	<b>53.79</b>	<b>82.46</b>	<b>91.73</b>	<b>92.00</b>	<b>88.27</b>

Table 6. Comparison of the proposed method and previous studies for helmetless head detection

Method	JSCI dataset			JSC2 dataset			database2 dataset		
	AP (%)	Pre (%)	Rec (%)	AP (%)	Pre (%)	Rec (%)	Acc (%)	Pre (%)	Rec (%)
[5]	43.41	53.18	68.15	40.47	52.27	78.59	88.26	89.18	82.64
[4]	52.45	57.50	68.15	54.30	50.00	78.13	91.37	91.61	94.04
[3]	70.16	67.53	77.04	<b>69.49</b>	<b>58.03</b>	<b>80.64</b>	87.49	87.22	92.75
<b>Proposed method</b>	<b>71.21</b>	<b>58.92</b>	<b>80.74</b>	66.63	53.79	82.46	<b>91.73</b>	<b>92.00</b>	<b>88.27</b>

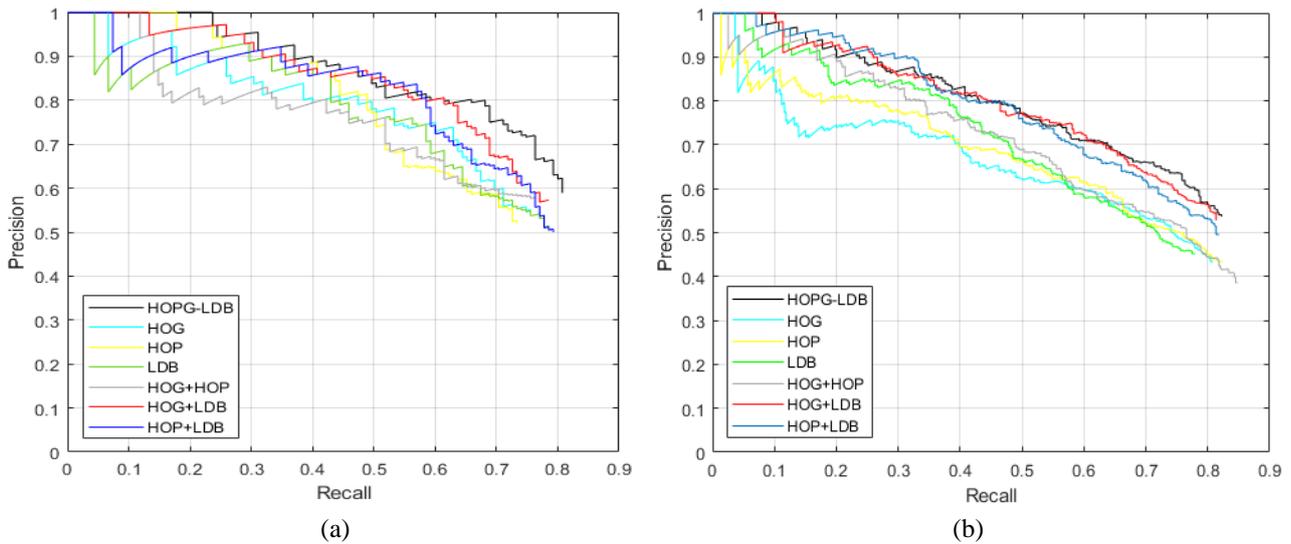


Figure. 14 Precision-recall curve of the descriptors: (a) *JSC1* dataset and (b) *JSC2* dataset

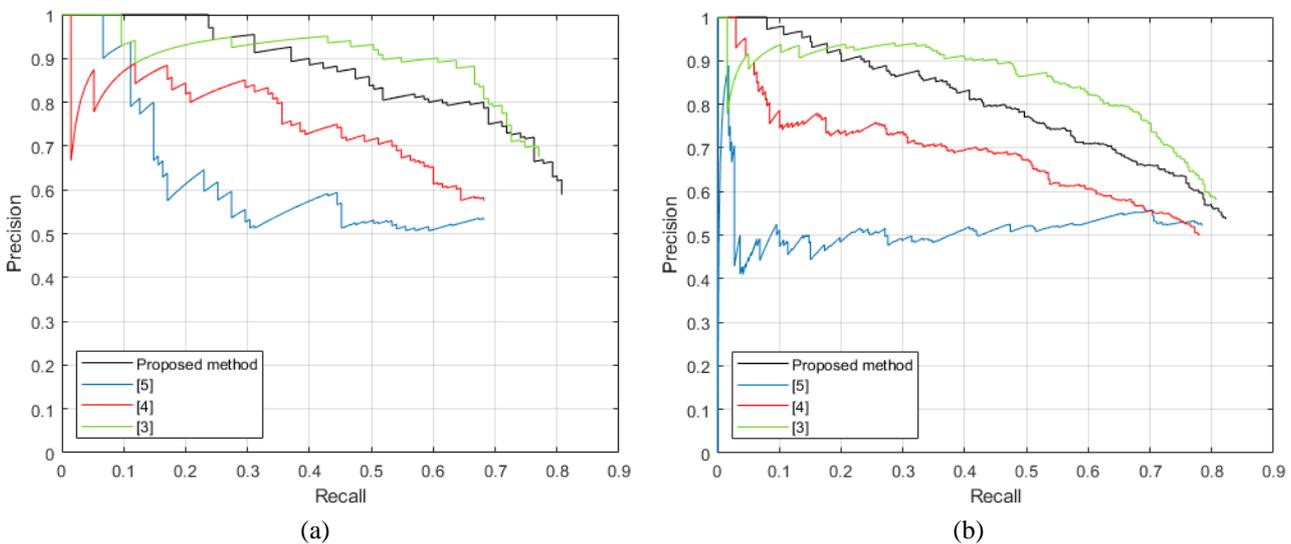


Figure. 15 Precision-recall curves of the proposed method and previous studies: (a) *JSC1* dataset and (b) *JSC2* dataset

## 5. Conclusion

This study proposes a method to detect motorcycle riders and passengers that do not wear helmets. Its main contribution is constructing a new descriptor, namely the HOPG-LDB, for motorcycle classification and helmetless head detection. The results are the proposed method capable of improving the performance of motorcycle classification compared to a single descriptor (HOG, HOP, and LDB) and combinations of two (HOG-HOP, HOG-LDB, and HOP-LDB). The proposed method has accuracies of 97.05% for the *JSC1*, 97.25% for *JSC2*, and 99.35% for *database1* datasets.

The proposed descriptor also improved the performance of helmetless head detection than the single and combinations of two descriptors. The proposed method is a type of hand-crafted feature

which is relatively suitable for use in real-time conditions. In addition, this method is suitable for datasets taken from rear-view and lighting variations. In future work, the feature selection process will be added to improve accuracy and reduce computing time.

## Conflicts of Interest

The authors declared no conflict of interest.

## Author Contributions

Conceptualization, Sutikno, Agus, and Afiahayati; methodology, Sutikno, Agus, and Afiahayati; software, Sutikno; validation, Sutikno, Agus, and Afiahayati; formal analysis, Sutikno, Agus, and Afiahayati; resources, Sutikno; data curation, Sutikno and Agus; writing—original draft preparation, Sutikno, Agus, and Afiahayati; writing—review and

editing, Sutikno, Agus, and Afiahayati; supervision, Agus and Afiahayati; funding acquisition, Agus.

## Acknowledgments

The authors are grateful to the Research Directorate of Universitas Gadjah Mada in the Rekognisi Tugas Akhir (RTA) 2021 scheme.

## References

- [1] W. H. O. (WHO), "Global Status Report on Road Safety 2018", Available at: <https://apps.who.int/iris/rest/bitstreams/1164010/retrieve>, 2021.
- [2] C. Huynh, T. Le, and K. Hamamoto, "Convolutional Neural Network for Motorbike Detection in Dense Traffic", In: *Proc. of the International Conf. on Communications and Electronics*, pp. 369-374, 2016.
- [3] L. Shine and C.V. Jiji, "Automated Detection of Helmet on Motorcyclists from Traffic Surveillance Videos: a Comparative Analysis using Hand-Crafted Features and CNN", *Multimedia Tools and Applications*, Vol. 79, pp. 14179-14199, 2020.
- [4] R. R. V. E. Silva, K. R. T. Aires, and R. M. S. Veras, "Detection of Helmets on Motorcyclists", *Multimedia Tools and Applications*, Vol. 77, No. 5, pp. 5659-5683, 2018.
- [5] K. Dahiya, D. Singh, and C. K. Mohan, "Automatic Detection of Bike-riders without Helmet using Surveillance Videos in Real-time", In: *Proc. of the International Joint Conf. on Neural Networks*, pp. 3046-3051, 2016.
- [6] H. K. Ragh and V. K. Asari, "Histogram of Oriented Phase and Gradient (HOPG) Descriptor for Improved Pedestrian Detection", In: *Proc. of the IS & T International Conf. on Electronic Imaging: Video Surveillance and Transportation Imaging Applications*, pp. 1-6, 2016.
- [7] H. Wang, D. Zhang, and Z. Miao, "Fusion of LDB and HOG for Face Recognition", In: *Proc. of the 37th Chinese Control Conf.*, pp. 9192-9196, 2018.
- [8] A. Leelasantitham and W. Wongsere, "Detection and Classification of Moving Thai Vehicles Based on Traffic Engineering Knowledge", In: *Proc. of International Conf. on ITS Telecommunication*, pp. 27-30, 2008.
- [9] S. Fazli, S. Muhammadi, and M. Rahmani, "Neural Network Based Vehicle Classification for Intelligent Traffic Control", *International Journal of Software Engineering & Applications*, Vol. 3, No. 3, pp. 17-22, 2012.
- [10] T. Marayatr and P. Kumhon, "Motorcyclist's Helmet Wearing Detection using Image Processing", *Advanced Materials Research*, Vol. 931-932, pp. 588-592, 2014.
- [11] Y. Dupuis, P. Subirats, and P. Vasseur, "Robust Image Segmentation for Overhead Real-time Motorbike Counting", In: *Proc. of International IEEE Conf. on Intelligent Transportation Systems*, pp. 3070-3075, 2014.
- [12] M. Ashvini, G. Revathi, B. Yogameena, and S. Saravanaperumaal, "View Invariant Motorcycle Detection for Helmet Wear Analysis in Intelligent Traffic Surveillance", In: *Proc. of International Conf. on Computer Vision and Image Processing*, pp. 175-185, 2016.
- [13] Z. Chen, T. Ellis, and S. A. Velastion, "Vehicle Detection, Tracking and Classification in Urban Traffic", In: *Proc. of International IEEE Conf. on Intelligent Transportation Systems*, pp. 951-956, 2012.
- [14] R. Silva, K. Aires, T. Santos, K. Abdala, R. Veras, and A. Soares, "Automatic Detection of Motorcyclists without Helmet", In: *Proc. of Latam America Computing Conf.*, pp. 1-7, 2013.
- [15] R. Waranusast, N. Bundon, and P. Pattanathaburt, "Machine Vision Techniques for Motorcycle Safety Helmet Detection", In: *Proc. of International Conf. on Image and Vision Computing New Zealand*, pp. 35-40, 2013.
- [16] C. Visnu, D. Singh, C. K. Mohan, and S. Babu, "Detection of Motorcyclists without Helmet in Videos using Convolutional Neural Network", In: *Proc. of International Joint Conf. on Neural Networks*, pp. 3036-3041, 2017.
- [17] A. S. Talautikar, S. Sanathanan, and C. N. Modi, "An Enhanced Approach for Detecting Helmet on Motorcyclists using Image Processing and Machine Learning Techniques", *Advanced Computing and Communication Technologies*, Vol. 702, pp. 109-119, 2019.
- [18] J. Mistry, A. K. Misraa, M. Agarwal, A. Vyas, V. M. Chudasama, and K. P. Upla, "An Automatic Detection of Helmeted and Non-helmeted Motorcyclist with License Plate Extraction using Convolutional Neural Network", In: *Proc. of Seventh International Conf. on Image Processing Theory, Tools and Applications*, pp. 1-6, 2017.
- [19] K. C. D. Raj, A. Chairat, V. Timtong, M. N. Dailey, and M. Ekpanyapong, "Helmet Violation Processing using Deep Learning", In: *Proc. of International Workshop on Advanced Image Technology*, pp. 1-4, 2018.
- [20] M. A. V. Forero, "Detection of Motorcycles and use of Safety Helmets with an Algorithm using

- Image Processing Techniques and Artificial Intelligence Models”, In: *Proc. of Joint Conf. for Urban Mobility in the Smart City*, pp. 1-9, 2018.
- [21] V. Mayya and A. Nayak, “Traffic Surveillance Video Summarization for Detecting Traffic Rules Violators using R-CNN”, In: *Proc. of International Conf. on Computer, Communication, and Computational Sciences*, pp. 117-126, 2017.
- [22] Y. Xu, J. Dong, B. Zhang, and D. Xu, “Background Modelling Methods in Video Analysis: A Review and Comparative Evaluation”, *CAAI Transactions on Intelligence Technology*, Vol. 1, No. 1, pp. 43-60, 2016.
- [23] N. Dalal and W. Triggs, “Histograms of Oriented Gradients for Human Detection”, In: *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, pp. 886-893, 2005.
- [24] P. Kovesi, “Phase Congruency Detects Corners and Edges”, In: *Proc. of the 7th Digital Image Computing: Techniques and Applications*, pp. 309-318, 2003.
- [25] X. Yuan and P. Shi, “Iris Feature Extraction using 2D Phase Congruency”, In: *Proc. of Third International Conf. on Information Technology and Applications*, pp. 437-441, 2005.
- [26] X. Yang and K. Cheng, “LDB: An Ultra-Fast Feature for Scalable Augmented Reality on Mobile Devices”, In: *Proc. of the IEEE International Symposium on Mixed and Augmented Reality*, pp. 49-57, 2012.
- [27] X. Chen, Z. Liu, and Z. Zhang, “The Measurement of Planning Surface Roughness by Neural Networks based on Image”, In: *Proc. of the Sixth International Conf. on Natural Computation*, pp. 705-708, 2010.
- [28] H. Mustafidah, S. Hartati, R. Wardoyo, and A. Harjoko, “Selection of Most Appropriate Backpropagation Training Algorithm in Data Pattern Recognition”, *International Journal of Computer Trends and Technology*, Vol. 14, No. 2, pp. 92-95, 2014.
- [29] R. Padilla, S. L. Netto, and E. A. B. D. Silva, “A Survey on Performance Metrics for Object-Detection Algorithms”, In: *Proc. of the International Conf. on Systems, Signals, and Image Processing*, pp. 237-242, 2020.