

Face Detection Using Haar Cascade Classifiers Based on Vertical Component Calibration

Cheol-Ho Choi¹, Junghwan Kim¹, Jongkil Hyun¹, Younghyeon Kim¹, and Byungin Moon^{1,2,*}

Abstract

The growing significance of the security and human management fields attracts active research related to face detection and recognition systems. Among these face detection techniques based on machine learning, Haar cascade classifiers are widely used because of their high accuracy for human frontal faces. However, the Haar cascade classifiers have a limitation in that the processing time increases as the number of false positives increases because they detect human faces based on the sub-window operation. Therefore, in this paper, a pre-processing method based on a 2D Haar discrete wavelet transform is proposed for face detection. The proposed method improves the processing speed by reducing the number of false positives through a vertical component calibration process using the vertical and horizontal components. The results of the face detection experiments that use a public test dataset comprising 2,845 images showed that the proposed method improved the processing speed by 32.05% and reduced the number of false positives by 25.46%, compared with those of the histogram equalization that shows the best performance case among conventional filter-based pre-processing methods. In addition, the performance of the proposed method is similar to those of conventional image contraction-based methods. In an experiment using a private dataset, the proposed method showed a 53.85% reduction in the total number of false positives compared with that of the Gaussian filter while maintaining the total number of true positives. The F_1 score of the proposed method shows a 1.39% improvement compared with those of Lanczos-3 that shows the best performance case.

Keywords

2D Haar Wavelet Transform, Haar Cascade Classifiers, Face Detection, Vertical Component Calibration

1. Introduction

As processor and chip technologies advance, various computer vision technologies for human-centered computing have been attracting considerable research attention. These technologies are being introduced in various fields such as Internet of Things (IoT), security, and autonomous driving [1–4]. Among computer vision technologies, face detection and recognition techniques are actively being studied because they can provide convenience to users in various domains, such as IoT environment-based security, management, and interpersonal communication [5–7]. In addition, owing to the coronavirus

* This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Corresponding Author: Byungin Moon (bihmoon@knu.ac.kr)

¹Graduate School of Electronic and Electrical Engineering, Kyungpook National University, Daegu, Korea

²School of Electronics Engineering, Kyungpook National University, Daegu, Korea

disease 2019 (COVID-19) outbreak, demand has increased for non-contact detection equipment and technology for biometric detection, such as body temperature and face detection [8, 9].

Face detection techniques can be based on machine learning or deep learning [10]. Deep learning-based face detection is generally based on neural networks [11, 12]. Zhu et al. [13] proposed a face detection method using a convolutional neural network (CNN) that utilizes single-stage headless face detection to overcome the limitations of computing power and storage. Guo et al. [14] proposed face detection using a CNN to improve processing speed. Although these studies have been conducted to improve the processing speed of deep learning, the many computational processes in the layers of the network architecture require a large processing time to compute the result [15]. Therefore, in such case, real-time processing is possible only when a high-performance processor and graphic processing unit (GPU) are used. For these reasons, most researchers focus on software because it requires significant resources to implement with digital logic.

In machine learning, which is a classical method in the field of artificial intelligence, cascade classifier architectures are typically used. These methods do not require high-performance processors or GPUs because the number of computations is smaller than those of the general deep learning-based methods. However, high accuracy or fast processing speed cannot be guaranteed. For these reasons, many studies are being conducted on the adaptive boosting (AdaBoost)-based Haar cascade classifiers, which were proposed by Viola and Jones [16, 17]. The Haar cascade classifiers have the advantages of a high detection rate for the human frontal face and an improvement in processing speed [10, 18]. Wu et al. [19] proposed a Euclidean distance to improve the detection accuracy. When the Euclidean distance, which is calculated by comparing the detected face feature with the trained face features, is lower than the threshold value, it is classified as a human frontal face. However, there is a drawback in that the processing time needed to calculate the Euclidean distance is increased because it requires a square root operation. Rishikeshan et al. [20] proposed morphological image processing to improve the detection accuracy. However, it has the drawback of slow processing speed because the proposed method includes a comparison step for checking the brightness, histogram equalization (HE), and morphological processing before the image is entered into the Haar cascade classifiers. Although recent studies have improved the accuracy, increased processing time still hinders real-time operation.

To improve the processing speed while maintaining the detection accuracy, this study proposes vertical component calibration, which preserves the appropriate edge information for face detection, based on 2D Haar discrete wavelet transform. The proposed method can reduce the number of false positives calibrating the zero-calibrated vertical and horizontal detail coefficient from the approximation detail coefficient. The number of false positives decrease because, if the vertical coefficient of a non-human frontal face is reduced, the reference value of the trained feature cannot be satisfied with a high probability. On the other hand, human frontal faces have a slight effect on the true positive rate because the vertical components are fewer than the horizontal components. The reduction of false positives improves the processing speed because unnecessary operations are reduced in the Haar cascade classifiers. In addition, the processing speed is improved because the input image size is also reduced by the 2D Haar discrete wavelet transform.

The remainder of this paper is organized as follows. Section 2 describes the Haar cascade classifiers and the 2D Haar discrete wavelet transform. The proposed method is described in Section 3, and the experimental results using the face detection dataset and benchmark (FDDB) [21], which is a public dataset, and a private test dataset are shown in Section 4. Finally, in Sections 5 and 6, the results of the study are discussed and conclusions are stated, respectively.

2. Background

2.1 Haar Cascade Classifiers

The Haar cascade classifiers, which use Haar-like features, were proposed by Viola and Jones [16, 17]. This method is widely used for object detection because of its simple structure, high detection rate, and fast detection speed; in particular, it exhibits excellent performance in human frontal face detection. The Haar-like feature calculates the feature value of the area through the difference in brightness values. Haar-like features classify the various features that exist on objects with different positions, size, and shapes. Fig. 1 presents examples of two-rectangle and three-rectangle shapes of Haar-like features used in the Haar cascade classifiers. When using these Haar-like features, it is possible to detect a specific object in an image. The human frontal face has features that can be used for classification (e.g., eyes, nose, and mouth). Therefore, a human frontal face can be detected by comparing calculated and trained feature value as reference values.

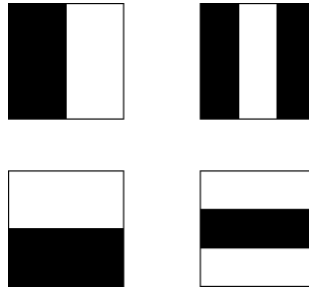


Fig. 1. Example of two-rectangle and three-rectangle shapes of Haar-like feature.

The feature value computation using a Haar-like feature is calculated by the difference in the sum of brightness value for dark and bright regions within a specific area. To obtain the sum of the brightness values, as many pixels as possible in the given area of the original image must be considered, and a significant amount of time is consumed in calculation. These problems occur because the calculation is based on the sub-window operation. To solve the problems, it is necessary to convert from the original image to an integral image before calculating the feature value. The integral image is generated by accumulating the pixel values of the original image in the lower-right direction. The integral image method is expressed mathematically as follows:

$$II(x_1, y_1) = \sum_{x < x_1} \sum_{y < y_1} I(x, y) \quad (1)$$

where $II(x_1, y_1)$ is the integral image, and $I(x, y)$ is the original input image. The sum of the brightness in a specific area using the integral image is obtained through the following equation:

$$S_{pixel} = P_{RB} - P_{RT} - P_{LB} + P_{LT} \quad (2)$$

where S_{pixel} is the pixel sum, P_{RB} is the right bottom value, P_{RT} is the right top value, P_{LB} is the left bottom value, and P_{LT} is the left top value of the area in the integral image. When using the two-rectangle Haar-like feature, the feature value of a specific area can be calculated using six coordinates of the integral image [22, 23].

In the Haar cascade classifiers, the classification results are determined by comparing the feature values with the trained values of the object. The Haar cascade classifiers consist of strong classifiers and weak classifiers. A strong classifier is a group of weak classifiers, generally called Haar-like features [24]. The strong classifier, which is in one of the classification stages, collects the comparison results of the weak classifiers included in the group and calculates the classification result of the relevant stage. Subsequently, it moves to the next stage when the result of classification in the current stage determines that the correct object has been identified. It is determined that the sub-window is the object to be detected

only when it passes through all strong classifier stages. If each strong classifier stage fails to pass, the sub-window is determined to be not the desired object area, and the operation for the sub-window is immediately terminated. It then moves to the next coordinate and begins the detection operation again.

Generally, Haar-like feature for classification are trained for windows that have a fixed size. This is called a sub-window, and it mostly uses a 20×20 or 24×24 size. The detection operation is performed by moving the sub-windows pixel by pixel in the original image for face detection operation. In the sub-window operation, it is difficult to detect all human frontal faces because of the fixed sub-window size. In other words, if the size of the sub-window or image is not changed, only a face of a specific size can be detected. To detect faces of various sizes, the image pyramid method is used to reduce the input image size so that the sub-window can detect human frontal faces. If several downscaled images are generated using the image pyramid method and then detection operations are performed on each of them, the faces of various sizes can be detected in the image with a fixed sub-window.

2.2 Discrete Wavelet Transform

The 2D discrete wavelet transform, used for image processing, is extended from the equation of 1D discrete wavelet transform used in earthquake analysis [25], electrocardiograms (ECGs) [26], and human vital sign detection [27]. The 1D discrete wavelet transform equation is expressed mathematically as follows [28]:

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot g[2n - k] \quad (3)$$

$$y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k] \cdot h[2n - k] \quad (4)$$

where $g[2n - k]$ is the scaling function, which is a low-pass filter, and $h[2n - k]$ is the wavelet function, which is a high-pass filter. The scaling and wavelet functions use mathematically predefined shapes according to the type of wavelet family [29]. Fig. 2 shows the approximation and detail coefficient computations using Equations (3) and (4). At each transformation level, the approximation coefficient (cA) is the low-frequency component, and the detail coefficient (cD) is the high-frequency component of the input signal $x[k]$. The results of the approximation and detail coefficient are half down-sampled because each transformation function shifts by $2n - k$ for computation.

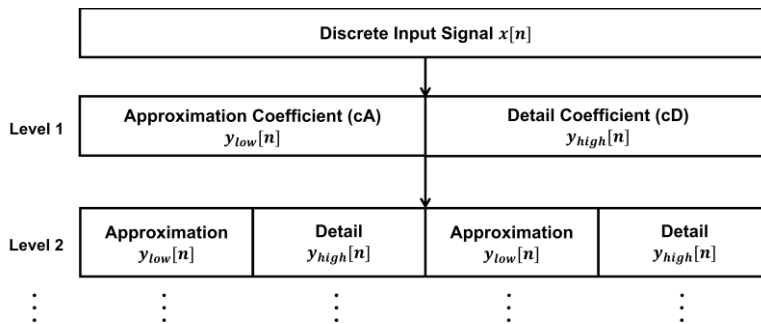


Fig. 2. Approximation and detail coefficient computation process using 1D discrete wavelet transform.

The 2D discrete wavelet transform uses the concept of 1D discrete wavelet transform to compute the related detail coefficients for image processing. The scaling and wavelet functions of the 2D discrete wavelet transform are expressed mathematically as follows [30]:

$$\varphi(x, y) = \varphi(x) \varphi(y) \quad (5)$$

$$\psi^H(x, y) = \varphi(x) \psi(y) \quad (6)$$

$$\psi^V(x, y) = \psi(x) \varphi(y) \quad (7)$$

$$\psi^D(x, y) = \psi(x) \psi(y) \quad (8)$$

where $\varphi(x, y)$ is the scaling function for the approximation detail coefficient; $\psi^H(x, y)$, $\psi^V(x, y)$, and $\psi^D(x, y)$ are the wavelet functions for horizontal, vertical, and diagonal detail coefficient, respectively. When the scaling and wavelet functions of the 2D discrete wavelet transform are separable, they can be expressed in the $f(x, y) = f_1(x)f_2(y)$ form, similar to the terms on the right side of Equations (5)–(8) [31]. In other words, the transformation functions to obtain coefficients in the 2D discrete wavelet transform can be divided into the scaling and wavelet function concepts of the 1D discrete wavelet transform. This can be computed sequentially by the transformation functions through operations in the row and column directions.

The approximation detail coefficient is computed using a scaling function for row and column directions, and the horizontal detail coefficient is computed by the wavelet function for the column direction. On the other hand, the diagonal detail coefficient is computed using the wavelet function for the row and column directions, and the vertical detail coefficient is computed by a scaling function for the column direction. The four types of detail coefficient results obtained through the 2D discrete wavelet transform are divided into the frequency domain channels of low-low (LL), low-high (LH), high-low (HL), and high-high (HH), respectively [32, 33].

3. Proposed Method

Haar cascade classifiers consists of weak and strong classifiers that generates a cascade structure of human frontal face detection based on sub-window operation. For this cascade structure, the processing time increases with an increase in the number of false positives. Therefore, various pre-processing methods are used to reduce the number of false positives. There are two types of conventional pre-processing methods: conventional filter-based and image contraction-based methods. In conventional filter-based methods, median, Gaussian filter, and HE are widely used to remove noise components and reduce the number of false positives. However, these methods still required a large amount of processing time and have a higher number of false positives, compared with image contraction-based methods. Conversely, the image contraction-based pre-processing methods have a higher processing speed because the image size and the number of false positives are reduced. However, when edge information suitable for face detection using Haar cascade classifiers is lost, the detection accuracy is decreased. A representative method in which edge information can be lost is a wavelet transform used to compute the approximation image. Meanwhile, when inappropriate edge information is included, the number of false positives is increased. That is, trade-off exists between the detection accuracy and the number of false positives depending on how much appropriate edge information is preserved [34, 35]. To reduce the number of false positives while maintaining the detection accuracy, the appropriate edge information needs to be preserved to satisfy the feature values for Haar cascade classifiers. Therefore, in this paper, we propose the vertical component calibration based on a 2D Haar discrete wavelet transform to preserve the appropriate edge information and remove the noise components to reduce the number of false positives while maintaining the detection accuracy.

Fig. 3 illustrates the entire face detection process using the Haar cascade classifiers with the proposed pre-processing method. The proposed method is a process of calibrating the vertical components of the image to preserve the appropriate edge information for human frontal face detection. To calibrate the vertical components, the desired image is generated by calibrating the vertical and horizontal detail

coefficient from the approximation detail coefficient. The desired image enters the strong classifier stage of Haar cascade classifiers as the input image, and the feature value is calculated using the sub-window operation. When the feature value of the sub-window satisfies all stages, the image is classified as a human frontal face. In the opposite case, the operation in the current sub-window is immediately terminated, and the same operation is performed by moving to the next pixel. When the sub-window operation for the input image is finished, the down-scaled images generated by the image pyramid method are sequentially entered. After the detection process for all image size is completed, multiple detected results for the same object are merged into a bounding box.

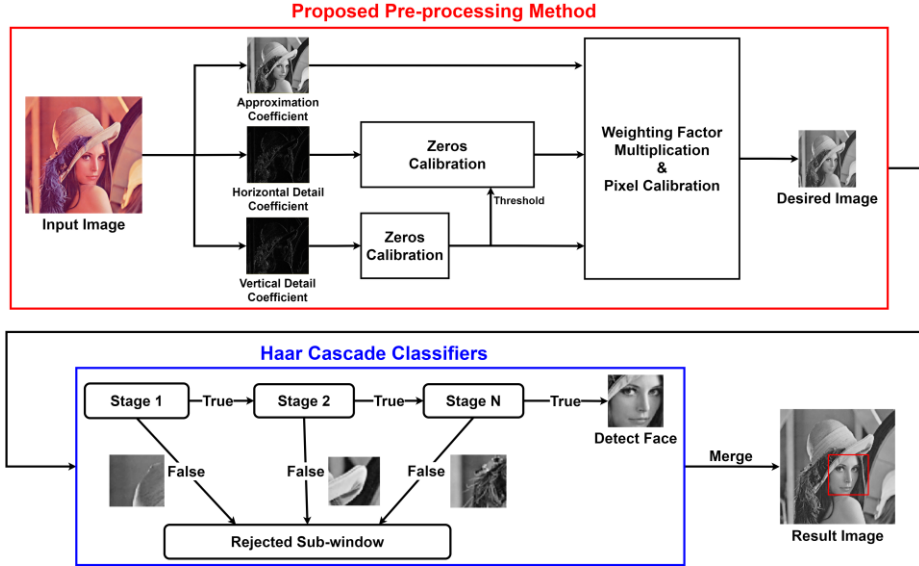


Fig. 3. Face detection process using the Haar cascade classifiers with the proposed method.

The proposed method aims to generate an image that, preserves the appropriate edge information by calibrating the vertical component, for Haar cascade classifiers. To calibrate the vertical component, the proposed method uses the three types of detail coefficient, which are called horizontal, vertical, and approximation detail coefficient, computed by the 2D Haar discrete wavelet transform. The approximation, vertical, and horizontal detail coefficient are mathematically expressed as follows:

$$x_{App}(n_1, n_2) = \sum_{i_1=0}^{K-1} \sum_{i_2=0}^{K-1} g(i_1) \cdot g(i_2) \cdot x(2n_1 - i_1, 2n_2 - i_2) \quad (9)$$

$$x_{Hori}(n_1, n_2) = \sum_{i_1=0}^{K-1} \sum_{i_2=0}^{K-1} g(i_1) \cdot h(i_2) \cdot x(2n_1 - i_1, 2n_2 - i_2) \quad (10)$$

$$x_{Vert}(n_1, n_2) = \sum_{i_1=0}^{K-1} \sum_{i_2=0}^{K-1} h(i_1) \cdot g(i_2) \cdot x(2n_1 - i_1, 2n_2 - i_2) \quad (11)$$

where K is the filter length of the transformation functions; $g(i_1)$ and $g(i_2)$ are scaling functions, which are the same as the low-pass filter; $h(i_1)$ and $h(i_2)$ are the wavelet function, which is the same as the high-pass filter; $x(2n_1 - i_1, 2n_2 - i_2)$ is the input image; $x_{App}(n_1, n_2)$ is the approximation detail coefficient; $x_{Hori}(n_1, n_2)$ is the horizontal detail coefficient; and $x_{Vert}(n_1, n_2)$ is the vertical detail coefficient, respectively. Fig. 4 shows the desired image-generation process based on Equations (9)–(11). In the 2D Haar discrete wavelet transform, the scaling function and wavelet function must satisfy the

orthogonal condition. In addition, the transformation function generated by using the scaling and wavelet function is in the form of a 2×2 matrix. After setting the components of the scaling and wavelet functions, the approximation detail coefficient can be obtained by the scaling function in the row and column directions. The horizontal detail coefficient can be obtained by the scaling function in the row direction and the wavelet function in the column direction. The vertical detail coefficient can be obtained using the wavelet function in the row direction and the scaling function in the column direction. Through a one-level transformation process, the vertical detail coefficient is calibrated to zero using the threshold value, whereas the horizontal detail coefficient is calibrated to zero using the zero-calibrated vertical detail coefficient as the threshold value. After the zero-calibration process, the desired image is generated by calibrating zero-calibrated vertical and horizontal detail coefficient, which are multiplied by the weighting factor, from the approximation detail coefficient.

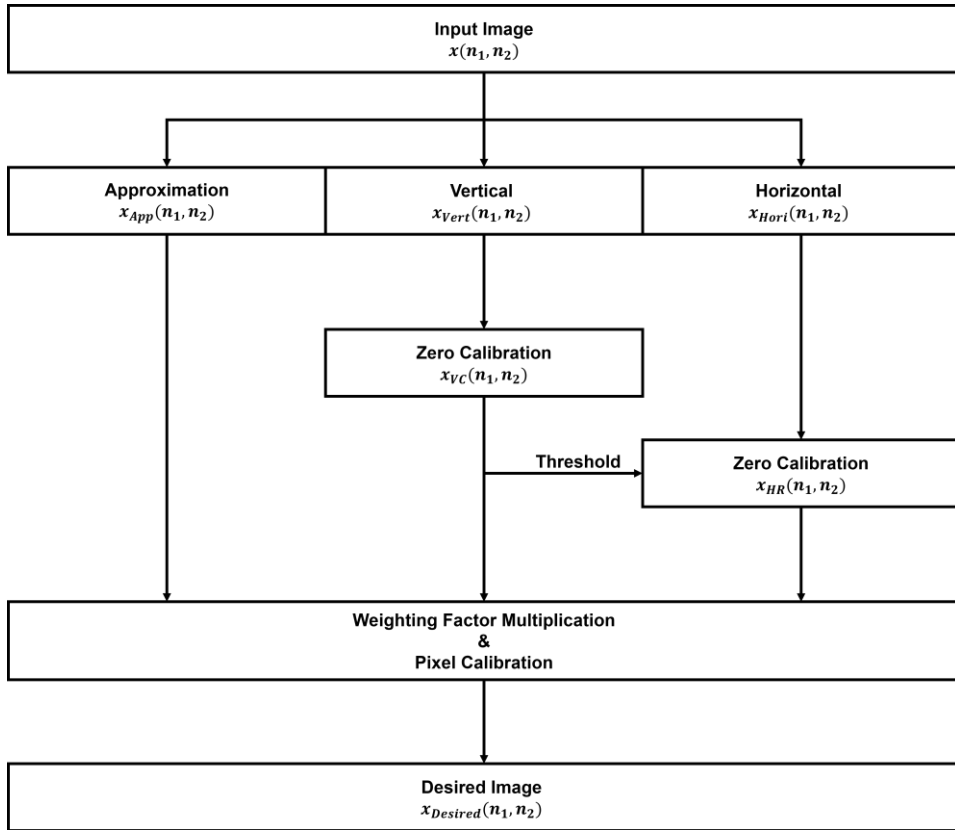


Fig. 4. Process of desired image generation using vertical component calibration for face detection using Haar cascade classifiers.

To generate the desired image for Haar cascade classifiers, the vertical and horizontal detail coefficient must be calibrated to zero before calibration from the approximation detail coefficient. The zero-calibrated vertical, horizontal detail coefficient, and the desired image are expressed mathematically as follows:

$$x_{VC}(n_1, n_2) = \begin{cases} x_{Vert}(n_1, n_2), & \text{for } x_{Vert}(n_1, n_2) \geq 0 \\ 0, & \text{for } x_{Vert}(n_1, n_2) < 0 \end{cases} \quad (12)$$

$$x_{HR}(n_1, n_2) = \begin{cases} x_{Hori}(n_1, n_2), & \text{for } x_{Hori}(n_1, n_2) \geq x_{VC}(n_1, n_2) \\ 0, & \text{for } x_{Hori}(n_1, n_2) < x_{VC}(n_1, n_2) \end{cases} \quad (13)$$

$$x_{Desired}(n_1, n_2) = x_{App}(n_1, n_2) + 2\alpha \times x_{VC}(n_1, n_2) - \alpha \times x_{HR}(n_1, n_2) \quad (14)$$

where $x_{VC}(n_1, n_2)$ is the zero-calibrated vertical detail coefficient, $x_{HR}(n_1, n_2)$ is the zero-calibrated horizontal detail coefficient, $x_{Desired}(n_1, n_2)$ is the desired image, that preserves the appropriate edge information, for frontal face detection using the Haar cascade classifiers, and α is weighting factor. In a grayscale image, the pixel value approaches zero as it becomes darker, and the pixel value approaches 255 as it becomes brighter. The vertical and horizontal detail coefficient can take both negative and positive values because the computation process involves the subtraction between the adjacent pixels at each coordinate. When the non-calibrated vertical detail coefficient is used, there is no difference in the accumulated value between the bright and dark regions of the Haar-like feature. In other words, the vertical calibration effect cannot be obtained in the human frontal face detection process using Haar-like features when a non-calibrated vertical detail coefficient is used. For this reason, the vertical detail coefficient is calibrated to zero before the calibration process, as shown in Equation (12), when it has a negative value. This is done because the face detection process is affected when extracting only the outer line of the vertical component. In addition, to compensate for the loss value of the pixel in the vertical component calibration, the horizontal detail coefficient is calibrated to zero, as shown in Equation (13), when it has a lower-than-zero-calibrated vertical detail coefficient. The reason for adjusting the value of the horizontal detail coefficient by using the vertical detail coefficient corrected to zero as the threshold value is to use the vertical component preferentially for the calibration process at the same coordinate in the image.

Based on Equation (14), the desired image is generated by calibrating the zero-calibrated vertical and horizontal detail coefficient, which are multiplied by the weighting factors, from the approximation detail coefficient. The vertical component is calibrated because the number of vertical components (e.g., nose) is less than the number of horizontal components (e.g., mouth, eyes) on the human frontal face. Therefore, the true positive rate of the original image can be maintained because only the outer line of the vertical component has a smaller effect on the original image when they are calibrated. Meanwhile, objects that are non-human frontal faces mostly have equal vertical and horizontal components or more vertical components than horizontal components. Due to characteristic of non-human face regions, the weighting factor for the vertical coefficient to reduce the number of false positives is twice that of the horizontal coefficient. Therefore, when the zero-calibrated vertical and horizontal coefficient are calibrated from the approximation detail coefficient, the number of false positives can be reduced while maintaining the detection accuracy.

4. Experimental Results

4.1 Public Dataset

A public dataset, FDDB [21], was used to verify the performance, which includes the true positive rate, processing time, and the number of false positives of Haar cascade classifiers with the proposed method, with the value of weighting factor α of 2. FDDB, consisting of 2,845 images with 5,171 faces, is a database with various poses, masks, and faces of various sizes. To evaluate the face detection performance of adopting the proposed method, we compared conventional filter-based methods (i.e., HE [36, 37], Gaussian [38, 39], and median filter [38, 39]) and image contraction-based methods (i.e., bicubic, Lanczos-2, Lanczos-3, and Haar discrete wavelet transform). In terms of the effect of vertical component calibration, the results indicate that Haar discrete wavelet transform only computes the denoising image, which is an approximation detail coefficient. For fair comparisons, we used the

CascadeObjectDetector built-in function of MATLAB R2021b (MathWorks, Natick, MA, USA) tool to detect bounding boxes of human frontal faces. For performance comparison, the xml file provided by open-source computer vision (OpenCV), was used with a scale factor of 1.2 for the image pyramid method and a 20×20 sub-window size. The haarcascade_frontalface_alt.xml file, provided by OpenCV, contains trained Haar-like feature information of the frontal face of humans. The indicators are discrete receiver operating characteristic (discROC) and continuous ROC (contROC), which are computed by using the evaluation method provided by Fddb. According to Fddb, the continuous and discrete scores for drawing the ROC curve are expressed mathematically as follows [21]:

$$S(d_i, I_j) = \frac{\text{area}(d_i) \cap \text{area}(I_j)}{\text{area}(d_i) \cup \text{area}(I_j)} \quad (15)$$

$$y_i = \delta_{S(d_i, v_i)} > 0.5 \quad (16)$$

$$y_i = S(d_i, v_i) \quad (17)$$

where d_i is the detection region, and I_j is the annotation region. Equation (16) is used to compute the discrete score for the discrete ROC curve, and Equation (17) is used to compute the continuous score for the continuous ROC curve. Fig. 5 shows the discrete ROC curves of the face detection results of adopting the proposed method and the conventional pre-processing methods. Fig. 6 shows the continuous ROC curves of the face detection results of adopting the proposed method and the conventional pre-processing methods. ROC curve is the graphical representation used to compare the performance of the method. The x-axis in Figs. 5 and 6 is the number of false positives, and the y-axis is the true positive rate. In this experimental result, the ROC curve using the Fddb evaluation method computes whether the detected bounding box before the merging step is true or false positive. When the area of the detected bounding box that overlaps with the ground truth is greater than the predefined threshold value, the number of false positives is fixed, and the true positive rate increases. Due to this computation process, when the number of false positives is small while having the value of similar true positive rate, the ROC curve converges to the final point value quickly. Therefore, the ROC curve for the proposed method exists at a higher position in the same region of the x-axis compared with conventional pre-processing methods, as shown in Figs. 6 and 7.

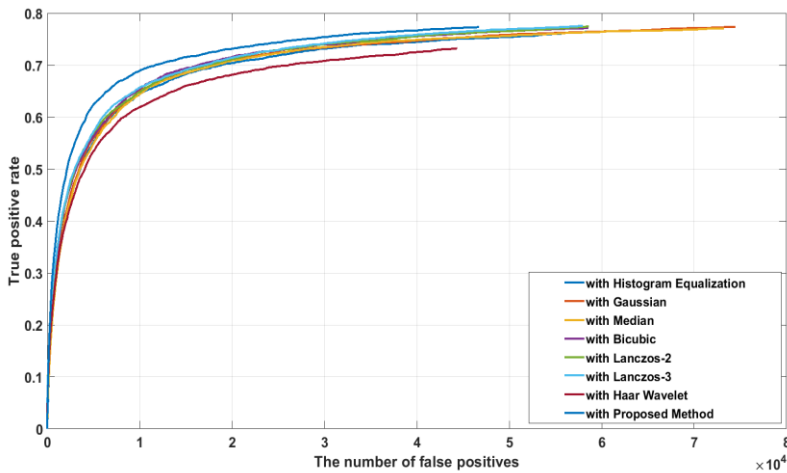


Fig. 5. Discrete ROC curves of face detection result of adopting the proposed method and the conventional pre-processing methods.

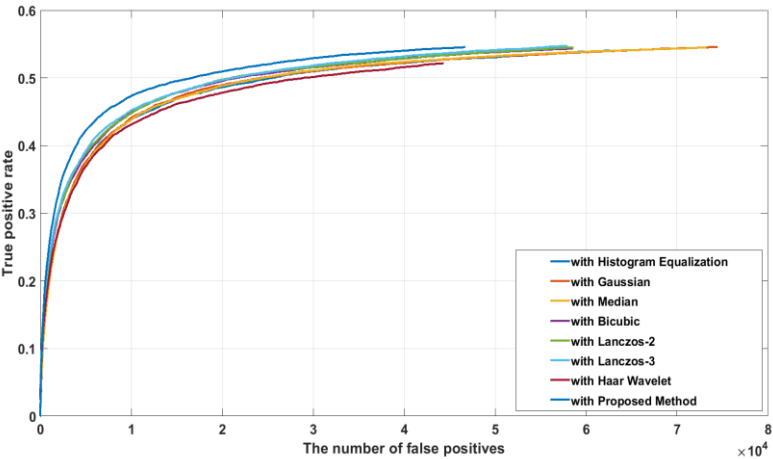


Fig. 6. Continuous ROC curves of face detection result of adopting the proposed method and the conventional pre-processing methods.

Table 1 presents the obtained values of the performance metrics, such as processing time, true positive rate at the final point of the ROC curve, and the number of false positives when adopting the proposed method and the conventional pre-processing methods. Among the traditional filter-based pre-processing methods, the HE pre-processing method obtained the best performance case in processing time and the number of false positives when using the haarcascade_frontalface_alt.xml file. When using the proposed method for Haar cascade classifiers, the processing time was 189.45 seconds, which was 32.05% faster than that of the HE method, and the number of false positives was 46,710, which was 25.46% less than that of the HE method. Among the image contraction-based methods, the Haar discrete wavelet transform, which only computes the approximation detail coefficient, shows the best performance in terms of the processing time and number of false positives. However, the true positive rate is decreased compared with the other image contraction-based methods. Although the processing time and number of false positives of the proposed method slightly increased compared with those of the Haar discrete wavelet transform, the true positive rate of the proposed method is similar to that of the other conventional image contraction-based methods. Overall, the proposed method is much better than the traditional filter-based method. In addition, the results show that the proposed method overcomes the trade-off between the number of false positives and true positive rate compare with the conventional image contraction-based methods.

Table 1. Performance of proposed method and conventional pre-processing methods using Fddb

	Processing time (s)	Number of false positives	Final point of true positive rate	
			CONTROC	DISCROC
Traditional filter-based METHOD				
With HE	278.8105	62661	0.540320	0.766196
With Gaussian	284.4435	74483	0.545753	0.773545
With median	282.1207	73222	0.544719	0.770837
Image contraction-based				
With bicubic	197.5510	58538	0.543523	0.771057
With Lanczos-2	197.5847	58611	0.545126	0.774250
With Lanczos-3	188.0214	57962	0.547588	0.775510
With Haar discrete wavelet transform	178.5786	44339	0.521732	0.732151
With proposed method	189.4536	46710	0.545502	0.773186

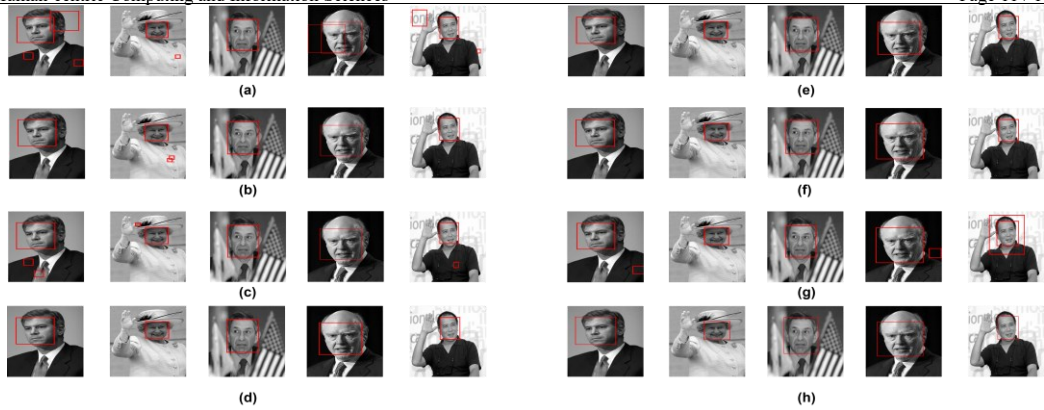


Fig. 7. Face detection results of Fddb test dataset after the merging step using the eight types of pre-processing methods: (a) HE, (b) Gaussian, (c) median, (d) bicubic, (e) Lanczos-2, (f) Lanczos-3, (g) Haar discrete wavelet transform, and (h) proposed method.

Fig. 7 shows the face detection results of sample images in the Fddb test dataset after the merging step when using the Haar cascade classifiers with the proposed method and the conventional pre-processing methods. Fig. 7(a)–7(h) depicts the results of face detection when using the HE, Gaussian filter, median filter, bicubic, Lanczos-2, Lanczos-3, Haar discrete wavelet transform, and proposed method, respectively. Fig. 7(a)–7(c) shows that adopting the conventional filter-based methods can detect human frontal face by removing noise. However, it can be confirmed that false positives exist in the non-human regions. Meanwhile, it can be visually confirmed that the number of false positives is reduced in the image contraction-based methods and the proposed method compared with the conventional filter-based methods, as shown in Fig. 7(d)–7(h).

4.2 Private Dataset

Figs. 8 and 9 show the performance of Haar cascade classifiers that adopted the proposed method and the conventional pre-processing methods, when applied to the private test dataset. For fair comparisons, we used with a scale factor of 1.2 and merge threshold factor of 1 for CascadeObjectDetector built-in function. The private test dataset consisted of 220 images with 794 faces in total of five image sizes. Evaluation results are classified as true positive when the intersection over union (IoU) [40] value about the annotation is 0.5 or more; otherwise, the results are classified as false positive. The total number of false positives that used the Gaussian filter was 247, showing the best performance among the conventional filter-based pre-processing methods, as shown in Fig. 8. When using the proposed pre-processing method, the total number of false positives was 114, which was 53.85% less than that of the Gaussian pre-processing method. The total number of true positives from adopting the proposed method was 658, which was similar to those of the conventional filter-based methods, as shown in Fig. 9. In image contraction-based methods, the total number of false positives that adopted the Haar discrete wavelet transform was 98, showing the best performance. Although the total number of false positives adopting the proposed method is slightly higher than that of Haar discrete wavelet transform, the total number of true positives adopting the proposed method shows better performance than that of Haar discrete wavelet transform, which shows the worst performance in terms of the total number of true positives.

For an objective evaluation, it is necessary to consider the precision, recall, and F_1 score, as well as the true positive (TP) and false positive (FP) results. The precision, recall, and F_1 score are expressed mathematically as follows [41, 42]:

$$Precision = \frac{TP}{TP + FP} \quad (18)$$

$$Recall = \frac{TP}{Total\ Faces} \quad (19)$$

$$F_1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (20)$$

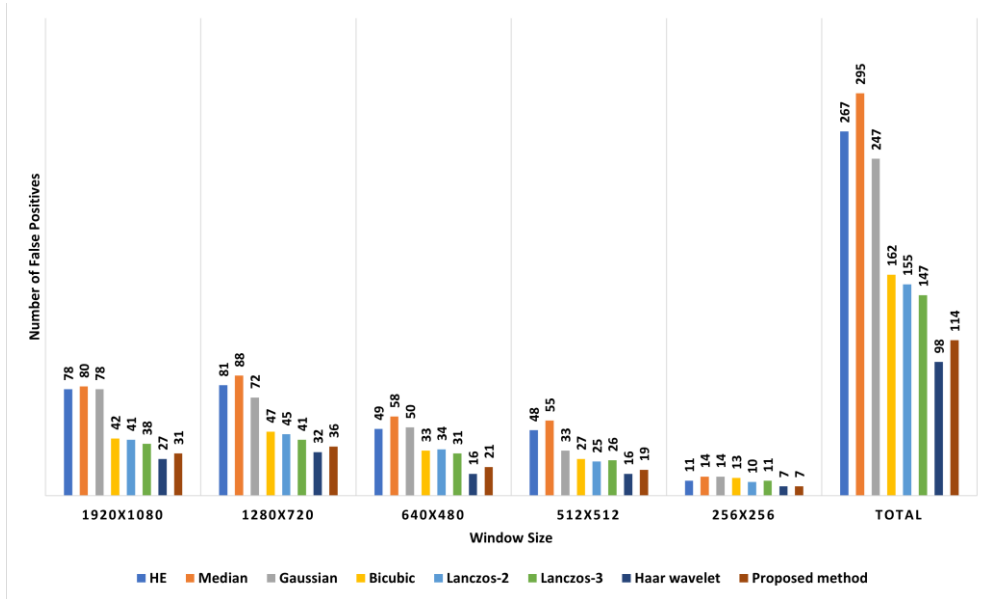


Fig. 8. Number of false positives adopting the proposed method and conventional methods for five image sizes of the private dataset.

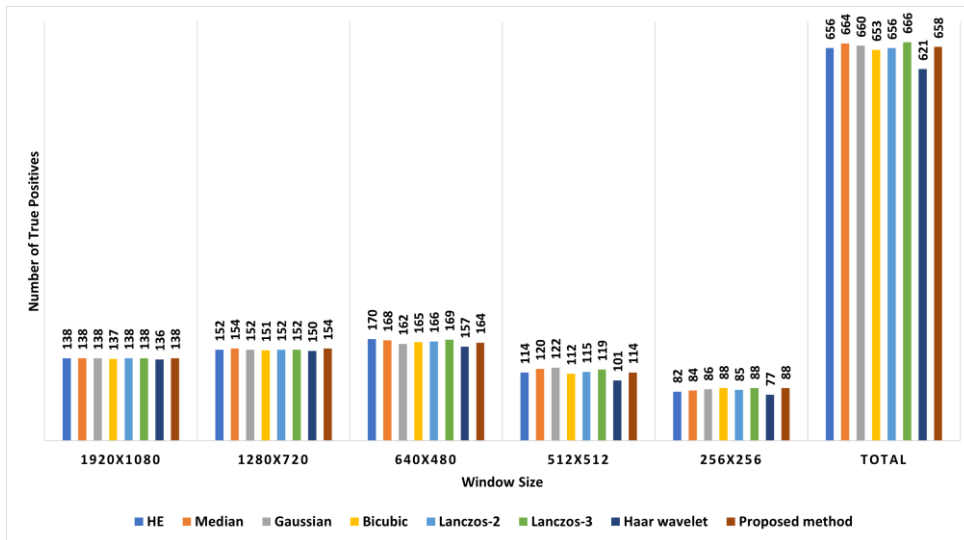


Fig. 9. Number of true positives adopting the proposed method and conventional methods for five image sizes of the private dataset.

The precision value, also called positive prediction, is defined as the ratio of true positives to all positives. Recall, which is widely called detection rate, is defined as the ratio of true positives to the total

number of faces. F_1 score is a mostly used member of the parametric family of F-measures, and it is defined as the harmonic mean of precision and recall [43]. Table 2 presents the results obtained from using Equations (18)–(20) to compare the objective performance using the private test dataset. In the precision factor, Gaussian filter obtained a value of 0.7277, which was the best performance case among the conventional filter-based pre-processing methods. When using the proposed method, the precision was 0.8525, which was improved by 17.15% compared with that of the Gaussian pre-processing method. In the precision factor of image contraction-based methods, Haar discrete wavelet transform obtained a value of 0.8637, which was the best performance. However, Haar discrete wavelet transform had the worst recall performance. The reason for the difference between precision and recall factor performance is as follows. When using Haar discrete wavelet transform, the total number of false positives has the best performance, as shown in Fig. 8. Therefore, the precision factor shows the best performance. However, the recall value was decreased because the total number of true positives showed the worst performance. The recall computed by adopting the proposed method has a value similar to that of the conventional pre-processing methods. In the F_1 score factor, the Gaussian and Lanczos-3 value were 0.7760 and 0.8289, which are the best performance case among the conventional filter-based and image contraction-based method, respectively. When using the proposed method, the F_1 score was 0.8404, which was improved by 8.30%, and 1.39% compared with the Gaussian, and Lanczos-3 pre-processing method, respectively. Overall, Table 2 shows that the face detection performance of the proposed method is improved compared with the conventional pre-processing method because the F_1 score of the proposed method is the highest.

Table 2. Performance of the proposed method and the conventional methods using the private dataset

	Precision	Recall	F_1 score
Traditional filter-based method			
With HE	0.7107	0.8262	0.7641
With Gaussian	0.7277	0.8312	0.7760
With median	0.6924	0.8363	0.7576
Image contraction-based method			
With Bicubic	0.8012	0.8224	0.8117
With Lanczos-2	0.8089	0.8262	0.8174
With Lanczos-3	0.8192	0.8388	0.8289
With Haar discrete wavelet transform	0.8637	0.7821	0.8209
With proposed method	0.8523	0.8287	0.8404

5. Discussion

In this study, face detection was performed using the proposed pre-processing method for Haar cascade classifiers. This study aimed to propose a method for improving the processing speed by reducing the number of false positives while maintaining the detection accuracy.

To evaluate the performance of Haar cascade classifiers using the proposed pre-processing method, we compared the proposed method with conventional filter-based and image contraction-based methods. In the conventional pre-processing methods, the filter-based methods are still limited in reducing the number of false positives and processing time. In contrast, the image contraction-based method has advantages in improving the processing speed by reducing the number of false positives. However, detection accuracy was decreased when appropriate edge information for face region was not preserved, as shown in result of Haar discrete wavelet transform. Conversely, when edge information for all area was preserved, the number of false positives increased, as shown in bicubic, Lanczos-2, and Lanczos-3 method. Thus, the conventional image contraction-based method has a trade-off between reducing the number of false positives and detection accuracy. To overcome the trade-off relationship, this paper

proposes vertical component calibration process to preserve the appropriate edge information for face region. The proposed method can reduce the number of false positives while maintaining the detection accuracy compared with the conventional filter-based and image contraction-based method. Therefore, the proposed method can be operated in real-time with high detection accuracy in various fields based on Haar cascade classifiers.

6. Conclusion

The face detection algorithm using the Haar cascade classifiers increases the processing time as the number of false positives increases. To improve the processing speed and reduce the number of false positives for face detection, this study proposed vertical component calibration process using a 2D Haar discrete wavelet transform for the Haar cascade classifiers. We evaluated and compared the performance using FDDB, which is a public test dataset consisting of 2,845 images. When using `haarcascade_frontalface_alt.xml` file, the face detection results of adopting the proposed method showed a 32.05% improvement in processing speed and 25.46% reduction in the number of false positives compared with those of the HE, which showed the best performance case among the conventional filter-based pre-processing methods. In addition, the processing time and detection accuracy of proposed method are similar to those of the conventional image contraction-based methods. In the private test dataset, the face detection results of adopting the proposed method showed a 53.85% reduction in the total number of false positives compared with that of the Gaussian pre-processing method, which showed the best performance case among the traditional filter-based pre-processing methods, while maintaining the total number of true positives. In addition, the value of F_1 factor of the proposed method, which considers both precision and recall, shows a 1.39% improvement compared with Lanczos-3, which shows the best performance among image contraction-based methods. The results computed using FDDB and private dataset show that the proposed method can overcome the trade-off between the number of false positives and detection accuracy. Therefore, the Haar cascade classifiers with the proposed method can be operated in real-time for various application, such as management and security for IoT based on face detection.

In a future work, the implementation and optimization of the proposed method in digital logic for face detection accelerator will be conducted based on the results of this study.

Acknowledgements

Not applicable.

Author's Contributions

Conceptualization, CHC, BM. Supervision, BM. Funding acquisition, BM. Methodology, CHC, JK, JH. Validation, CHC, JK. Data Curation, CHC, JH, YK. Writing of original draft, CHC, YK, BM. Writing of the review and editing, CHC, JK, JH, BM. Software, CHC. Visualization, CHC. Formal analysis, CHC.

Funding

This research was supported by the Multi-Ministry Collaborative R&D program (R&D program for complex cognitive technology) through the National Research Foundation of Korea (NRF) funded by Ministry of Trade, Industry and Energy (No. NRF-2018M3E3A1057248).

Competing Interests

The authors declare that they have no competing interests.

References

- [1] S. Pawar, V. Kithani, S. Ahuja, and S. Sahu, "Smart home security using IoT and face recognition," in *Proceedings of 2018 4th International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1-6.
- [2] N. Mostakim, R. R. Sarkar, and M. A. Hossain, "Smart locker: IoT based intelligent locker with password protection and face detection approach," *International Journal of Wireless and Microwave Technologies*, vol. 9, no. 3, pp. 1-10, 2019.
- [3] A. Zaarane, I. Slimani, W. Al Okaishi, I. Atouf, and A. Hamdoun, "Distance measurement system for autonomous vehicles using stereo camera," *Array*, vol. 5, article no. 100016, 2020. <https://doi.org/10.1016/j.array.2020.100016>
- [4] M. Wen, J. Park, and K. Cho, "A scenario generation pipeline for autonomous vehicle simulators," *Human-centric Computing and Information Sciences*, vol. 10, article no. 24, 2020. <https://doi.org/10.1186/s13673-020-00231-z>
- [5] J. Zhu, F. Yu, G. Liu, M. Sun, D. Zhao, Q. Geng, and J. Su, "Classroom roll-call system based on ResNet networks," *Journal of Information Processing Systems*, vol. 16, no. 5, pp. 1145-1157, 2020.
- [6] H. Y. Suen, K. E. Hung, and C. L. Lin, "Intelligent video interview agent used to predict communication skill and perceived personality traits," *Human-centric Computing and Information Sciences*, vol. 10, article no. 3, 2020. <https://doi.org/10.1186/s13673-020-0208-3>
- [7] I. S. Na, C. Tran, D. Nguyen, and S. Dinh, "Facial UV map completion for pose-invariant face recognition: a novel adversarial approach based on coupled attention residual UNets," *Human-centric Computing and Information Sciences*, vol. 10, article no. 45, 2020. <https://doi.org/10.1186/s13673-020-00250-w>
- [8] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: a novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," *Sustainable Cities and Society*, vol. 65, article no. 102600, 2021. <https://doi.org/10.1016/j.scs.2020.102600>
- [9] M. N. Mohammed, H. Syamsudin, S. Al-Zubaidi, R. Ramli, and E. Yusuf, "Novel COVID-19 detection and diagnosis system using IOT based smart helmet," *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 7, pp. 2296-2303, 2020.
- [10] A. Srivastava, S. Mane, A. Shah, N. Shrivastava, and B. Thakare, "A survey of face detection algorithms," in *Proceedings of 2017 International Conference on Inventive Systems and Control (ICISC)*, Coimbatore, India, 2017, pp. 1-4.
- [11] B. Peng and A. K. Gopalakrishnan, "A face detection framework based on deep cascaded full convolutional neural networks," in *Proceedings of 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, Singapore, 2019, pp. 47-51.
- [12] K. Smelyakov, A. Chupryna, O. Bohomolov, and I. Ruban, "The neural network technologies effectiveness for face detection," in *Proceedings of 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*, Lviv, Ukraine, 2020, pp. 201-205.
- [13] L. Zhu, F. Chen, and C. Gao, "Improvement of face detection algorithm based on lightweight convolutional neural network," in *Proceedings of 2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2020, pp. 1191-1197.
- [14] G. Guo, H. Wang, Y. Yan, J. Zheng, and B. Li, "A fast face detection method via convolutional neural network," *Neurocomputing*, vol. 395, pp. 128-137, 2020.
- [15] Y. LeCun, "1.1 deep learning hardware: past, present, and future," in *Proceedings of 2019 IEEE International Solid-State Circuits Conference-(ISSCC)*, San Francisco, CA, 2019, pp. 12-19.
- [16] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, HI, 2001, pp. 511-518.
- [17] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.
- [18] R. Vij and B. Kaushik, "A survey on various face detecting and tracking techniques in video sequences," in *Proceedings of 2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, Madurai, India, 2019, pp. 69-73.

- [19] H. Wu, Y. Cao, H. Wei, and Z. Tian, "Face recognition based on Haar like and Euclidean distance," *Journal of Physics: Conference Series*, vol. 1813, article no. 012036, 2021. <https://doi.org/10.1088/1742-6596/1813/1/012036>
- [20] C. A. Rishikeshan, C. Rajesh Kumar Reddy, and M. K. V. Nandimandalam, "An improved approach for face detection," in *Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*. Singapore: Springer, 2021, pp. 811-816.
- [21] V. Jain and E. Learned-Miller, "FDDB: a benchmark for face detection in unconstrained settings," University of Massachusetts, Amherst, MA, *Technical Report No. UMCS-2010-009*, 2010.
- [22] M. G. Krishna and A. Srinivasulu, "Face detection system on AdaBoost algorithm using Haar classifiers," *International Journal of Modern Engineering Research*, vol. 2, no. 5, pp. 3556-3560, 2012.
- [23] D. Kim, J. Hyun, and B. Moon, "Memory-efficient architecture for contrast enhancement and integral image computation," in *Proceedings of 2020 International Conference on Electronics, Information, and Communication (ICEIC)*, Barcelona, Spain, 2020, pp. 1-4.
- [24] C. Zhao, P. Wang, J. Chen, and W. Yang, "A weak moving point target detection method based on high frame rate SAR image sequences and machine learning," in *Proceedings of 2020 IEEE International Geoscience and Remote Sensing Symposium*, Waikoloa, HI, 2020, pp. 2795-2798.
- [25] A. Heidari and N. Majidi, "Earthquake acceleration analysis using wavelet method," *Earthquake Engineering and Engineering Vibration*, vol. 20, no. 1, pp. 113-126, 2021.
- [26] D. Zhang, S. Wang, F. Li, J. Wang, A. K. Sangaiah, V. S. Sheng, and X. Ding, "An ECG signal de-noising approach based on wavelet energy and sub-band smoothing filter," *Applied Sciences*, vol. 9, no. 22, article no. 4968, 2019. <https://doi.org/10.3390/app9224968>
- [27] E. L. Chuma and Y. Iano, "A movement detection system using continuous-wave Doppler radar sensor and convolutional neural network to detect cough and other gestures," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 2921-2928, 2020.
- [28] C. H. Choi, J. H. Park, H. N. Lee, and J. R. Yang, "Heartbeat detection using a Doppler radar sensor based on the scaling function of wavelet transform," *Microwave and Optical Technology Letters*, vol. 61, no. 7, pp. 1792-1796, 2019.
- [29] C. U. Kumari, A. S. D. Murthy, B. L. Prasanna, M. P. P. Reddy, and A. K. Panigrahy, "An automated detection of heart arrhythmias using machine learning technique: SVM," *Materials Today: Proceedings*, vol. 45, pp. 1393-1398, 2021.
- [30] R. C. Gonzales, and R. E. Woods, *Digital Image Processing*, 4th ed. New York, NY: Pearson, 2018.
- [31] C. L. Liu, "A tutorial of the wavelet transform," 2010 [Online]. Available: <http://disp.ee.ntu.edu.tw/tutorial/WaveletTutorial.pdf>.
- [32] P. S. Tsai and T. Acharya, "Image up-sampling using discrete wavelet transform," in *Proceedings of the 2006 Joint Conference on Information Sciences (JCIS)*, Kaohsiung, Taiwan, 2006.
- [33] M. A. Gungor, "A comparative study on wavelet denoising for high noisy CT images of COVID-19 disease," *Optik*, vol. 235, article no. 166652, 2021. <https://doi.org/10.1016/j.ijleo.2021.166652>
- [34] M. U. Yaseen, A. Anjum, O. Rana, and R. Hill, "Cloud-based scalable object detection and classification in video streams," *Future Generation Computer Systems*, vol. 80, pp. 286-298, 2018.
- [35] M. A. Zulkhairi, Y. M. Mustafah, Z. Z. Abidin, H. F. M. Zaki, and H. A. Rahman, "Car detection using cascade classifier on embedded platform," in *Proceedings of 2019 7th International Conference on Mechatronics Engineering (ICOM)*, Putrajaya, Malaysia, 2019, pp. 1-3.
- [36] K. Padmaja and T. N. Prabakar, "FPGA based real time face detection using Adaboost and histogram equalization," in *Proceedings of IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM)*, Nagapattinam, India, 2012, pp. 111-115.
- [37] S. M. Bah and F. Ming, "An improved face recognition algorithm and its application in attendance management system," *Array*, vol. 5, article no. 100014, 2020. <https://doi.org/10.1016/j.array.2019.100014>
- [38] P. Mazurek and T. Hachaj, "Robustness of Haar feature-based cascade classifier for face detection under presence of image distortions," In *Image Processing and Communications*. Cham, Switzerland: Springer, 2019, pp. 14-21.
- [39] L. T. H. Phuc, H. Jeon, N. T. N. Truong, and J. J. Hak, "Applying the Haar-cascade algorithm for detecting safety equipment in safety management systems for multiple working environments," *Electronics*, vol. 8, no. 10, article no. 1079, 2019. <https://doi.org/10.3390/electronics8101079>

- [40] H. Rezaatofghi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: a metric and a loss for bounding box regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 2019, pp. 658-666.
- [41] A. Kumar, M. Kumar, and A. Kaur, "Face detection in still images under occlusion and non-uniform illumination," *Multimedia Tools and Applications*, vol. 80, no. 10, pp. 14565-14590, 2021.
- [42] H. Shi, X. Chen, and M. Guo, "Re-SSS: rebalancing imbalanced data using safe sample screening," *Journal of Information Processing Systems*, vol. 17, no. 1, pp. 89-106, 2021.
- [43] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, article no. 6, 2020. <https://doi.org/10.1186/s12864-019-6413-7>