Fast Object Detection Algorithm using Edge-based Operation Skip Scheme with Viola-Jones Method

Cheol-Ho Choi*, Joonhwan Han, Jeongwoo Cha, Jungho Shin, Hyun Woo Oh
Pangyo Research and Development (R&D) Center
Hanwha Systems, Co., Ltd.
Seongnam, Republic of Korea
*cheoro1994@hanwha.com

***Abstract*—Machine learning approaches are preferred over deep learning in embedded systems due to their resource efficiency. The widely adopted Viola-Jones method and related algorithms are selected for their high detection accuracy and reasonable processing speed. However, a limitation arises as processing time increases with additional classification iterations based on sub-window operations. To address this issue, we propose an enhanced object detection algorithm that incorporates the Viola-Jones method with edge component calibration and an edge-based operation skip scheme. The introduction of edge component calibration ensures detection performance comparable to conventional methods. This scheme, relying on edge values, significantly reduces unnecessary computations in the background, leading to a marked decrease in classification operations compared to conventional methods. Visual comparisons in experimental results demonstrate that our method increases the detection precision factor while maintaining recall. In terms of classification operations, our approach reduces their number by 31.38% to 85.78% compared to conventional methods. In simpler terms, our method improves processing speed by minimizing classification operations, making it well-suited for embedded systems with limited resource utilization.**

***Index Terms*—Object detection, Machine learning, Viola-Jones method, Skip scheme, Wavelet transform**

I. Introduction

The rapid advancement of chip performance has led to the widespread application of computer vision technologies in various industry sectors, including autonomous driving and monitoring systems [1], [2]. However, practical implementations are constrained by associated production costs, limiting the availability of algorithms for commercialized platforms. Notably, within in-cabin systems, the european new car assessment programme (Euro NCAP) document recommends (in effect, mandates) the inclusion of Child Presence Monitoring (CPD) systems in vehicles by 2025. To meet this requirement, manufacturers are exploring the deployment of low-cost sensors.

Facilitating the implementation of in-cabin systems using object recognition often involves integrating both machine learning and deep learning techniques [3]. This combination enhances processing speed by dividing the object detection and recognition steps. Machine learning techniques are preferred for object detection due to their fast processing speed and simple architecture compared to deep learning techniques.

Among machine learning-based object detection system, the traditional Viola-Jones method stands out for its simplicity, utilizing pre-trained Haar-like features, resulting in fast processing speeds [4]–[7]. However, drawbacks arise as processing time increases with higher image resolution and an augmented number of classification iterations [8]. These limitations are crucial for real-time object detection in resource limited embedded systems, such as low-cost field-programmable gate array (FPGA) environments. To address these challenges, ongoing research endeavors are actively exploring potential solutions.

Hyun et al. proposed a skip scheme based on the Intersection over Union (IoU) for the Viola-Jones method [9]. While effective in speeding up processing by skipping classification operations for the same object region, its effectiveness diminishes as the number of objects in the input image decreases. Subsequently, Choi et al. presented a wavelet transform-based edge component calibration method for the Viola-Jones method, aiming to improve processing speed while maintaining detection accuracy [8]. This method utilizes two-dimensional (2-D) Haar wavelet transform, reducing image resolution to enhance processing speed. However, while Choi's method significantly improves processing speed, it has the drawback of not ensuring a consistently fast processing speed due to the continued large number of operation repetitions for classification in the background area.

To overcome the identified limitations in these follow-up studies, this paper proposes a novel object detection method adopting an edge-based skip scheme and calibration method. In this method, the 2-D Haar wavelet analysis is introduced for two operations: 1) edge-based operation skip scheme and 2) edge component calibration.

- Edge-based Operation Skip Scheme: In this scheme, the classification operation is skipped when edge components have invalid values, and it is performed when edge components have valid values
- Edge Component Calibration: To maintain the performance factors of precision and recall, we incorporate the revised edge component calibration method introduced by Choi in our proposed approach

In our proposed method, we can maintain precision and recall factors while improving the processing speed for object detection operations by incorporating these two concepts into the Viola-Jones method.

II. Proposed Method

Fig. 1 illustrates the operational process of the proposed object detection method. In an effort to enhance processing speed and detection accuracy, we have introduced the edge-based skip scheme and edge component calibration method.

A. Edge Component Calibration

For the computation of edge components, our proposed object detection method incorporates 2-D Haar wavelet analysis transform as a pre-processing step. The formulas for the wavelet transform are as follows:

$$I_A(x,y) = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1} I_{in}(x,y)T_A(x,y) \tag{1}$$

$$I_H(x,y) = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1} I_{in}(x,y)T_H(x,y) \tag{2}$$

$$I_V(x,y) = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1} I_{in}(x,y)T_V(x,y) \tag{3}$$

$$I_D(x,y) = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1} I_{in}(x,y)T_D(x,y) \tag{4}$$

Here, $I_{in}(x,y)$ represents the input image; $I_A(x,y)$, $I_H(x,y)$, $I_V(x,y)$ and $I_D(x,y)$ correspond to the approximation, horizontal, vertical, and diagonal edge components, respectively. Additionally, $T_A(x,y)$, $T_H(x,y)$, $T_V(x,y)$ and $T_D(x,y)$ denote the transformation functions for computing the approximation, horizontal, vertical, and diagonal components, respectively. Regarding the approximation component, $I_A(x,y)$ signifies the de-noised image with high-frequency and noise components removed. In summary, the de-noised image and three types of edge components for the calibration process are computed using (1)–(4).

After the wavelet analysis process, the resulting de-noised image and three edge components undergo the calibration process. This calibration process is designed to enhance the detection accuracy for objects. Within the edge calibration process, the three edge components are adjusted to match the de-noised image. The formula for the calibration process is as follows:

$$R(x') = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{5}$$

$$I_M(x,y) = R(I_H(x,y)) + R(I_V(x,y)) - R(I_D(x,y)) \tag{6}$$

$$I_{out}(x,y) = I_A(x,y) + I_M(x,y) \tag{7}$$

Here, $R(x')$ represents the operation function for the rectified linear unit (ReLU). $I_M(x,y)$ denotes the merged edge component formed using the horizontal, vertical, and diagonal components. $I_{out}(x,y)$ is the output image obtained after calibrating the merged edge component. Prior to the calibration process, the ReLU operation is applied to preserve valid values for each edge component. Following the ReLU operation, the calibration operation takes place. In the context of edge component calibration, we introduce the operational concept from the method proposed by Choi [8]. The distinctive feature of our proposed calibration method is the inclusion of a subtraction process utilizing the diagonal component. By incorporating this subtraction process with the diagonal component, we can suppress over-calibrated values resulting from the horizontal and vertical components.

B. Edge-based Operation Skip Scheme

Through the calibration step, the merged pyramidal edge component and de-noised pyramidal image are computed because of N-level wavelet transform. The pyramidal de-noised image serves as the input for the classification process. Our proposed object detection method employs the Viola-Jones method for classification, known for its high detection accuracy among various machine learning techniques. However, processing time significantly increases as input image resolution rises, as the classification operation is executed for all coordinates of the input image. To address this, we introduce a skip scheme that utilizes the merged edge component to enhance processing speed. As depicted in Fig. 1, the classification operation is skipped when the merged edge component value is invalid. Conversely, the classification operation is only performed when the merged edge component value is valid. This skip scheme concept improves processing speed.

Fig. 2 illustrates the calculation process for selecting the reference coordinate in the skip scheme. To perform this calculation, predefined values for α and β must be set for the sub-window of $I_M(x,y)$. The formula for selecting the reference coordinate is as follows:

$$(x_{Ref}, y_{Ref}) = (N \times \alpha, M \times \beta) \tag{8}$$

The x_{Ref} and y_{Ref} represent the reference coordinates for the vertical and horizontal directions, respectively. N and M denote the horizontal and vertical window sizes of the sub-window. β is the margin factor for calculating the horizontal coordinate, and α is the margin factor for calculating the vertical coordinate. As shown in (8), the margin factors of α and β are set from zero to one. Fig. 2(b) illustrates a scenario where the margin factors of α and β are set to 0.5 and 0.2, respectively. In this case, the reference coordinate for the skip scheme is (10, 4) when using the 20 $\times$ 20 window size. Consequently, the classification operation is executed when the pixel value of the merged edge component located at (10, 4) is a positive value. Conversely, the classification operation is skipped when the pixel value of the merged edge component located at (10, 4) is a negative value. Through the proposed edge calibration and classification method, the output image containing the detected objects can be computed.

III. Experimental Results

To conduct quantitative and qualitative evaluations and comparisons based on the experimental results, we used three images: 1) Lena and 2) Solvay Conference 1927.

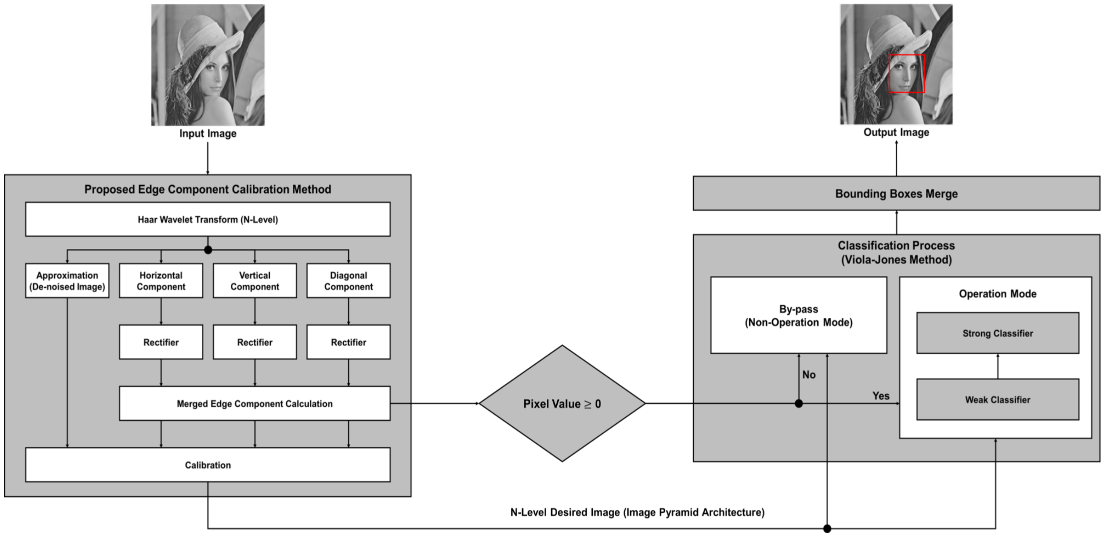

Fig. 1. Operation process of the proposed method: Edge-based operation skip scheme and edge component calibration in Viola-Jones method.

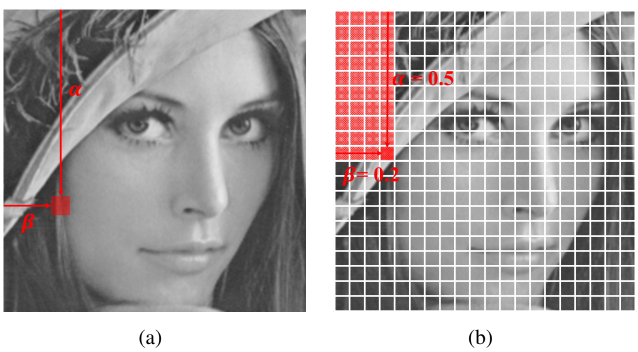

Fig. 2. Example of selecting the reference coordinate in the proposed operation skip scheme.

To ensure a fair experiment, we employed the 'haarcascade_frontalface_alt.xml' file provided by the open-source computer vision (OpenCV) library. In the 'haarcascade_frontalface_alt.xml', the square window size and strong classifier stage are set to 20 and 22, respectively. Regarding operational factors, we configured the IoU, scale factor, and margin factors α and β to 0.5, 2, 0.5, and 0.05, respectively.

A. Quantitative Evaluation (Visual Confirmation)

Fig. 3–4 display the experimental results using both conventional and proposed methods for the images of 'Lena' and 'Solvay Conference 1927'.

In Fig. 3(a) and 4(a), the traditional Viola-Jones method exhibits precise detection of target objects (faces) in each image. However, it also produces incorrect detection results in the background area. Fig. 3(b) and 4(b) show the results using the method proposed by Hyun [9], which, similar to the experimental results of the traditional Viola-Jones method, demonstrates correct target detection results but retains incorrect detection results in the background. This similarity arises because the method proposed by Hyun [9] primarily skips the detection process based on IoU, resulting in a pattern similar to Fig. 3(a) and 4(a) without significant differences. In Fig. 3(c) and 4(c), the results using the method proposed by Choi [8] reveal correct detection results for all targets, with a noticeable reduction in incorrect detection results in the background compared to the traditional Viola-Jones method and Hyun's method.

Fig. 3(d) and 4(d) illustrate the outcomes of the proposed method, showcasing accurate target detection results and a substantial reduction in incorrect detection results compared to the experimental results of conventional methods. In other words, the proposed method effectively minimizes false positives in comparison to conventional methods. Consequently, when evaluating precision-recall performance for the results in Fig. 3–4, our proposed method demonstrates high precision and equivalent recall performance. Precision is influenced by the number of false positives, making lower false positives indicative of higher precision. Conversely, the recall value increases with the number of detected targets and is independent of false positives.

B. Qualitative Evaluation

To conduct a qualitative comparison and evaluation, we analyze the required number of classification iterations using both the proposed and conventional methods. Table I displays the necessary number of classification iterations when employing both the proposed and conventional methods with one weak classifier for three images: 'Lena' and 'Solvay Conference 1927' as presented in Fig. 3–4.

- Lena image: Gray-scale, 8-bit, and 512 $\times$ 512 resolution.
- Solvay Conference 1927 image: Gray-scale, 8-bit, and 1280 $\times$ 886 resolution.

As shown in Table I, when applying the traditional Viola-Jones method to the 'Lena' image, it requires $15.50M$ classification iterations for object detection. In contrast, the method

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

Fig. 3. Experimental result using the ‘Lena’ image: (a) traditional Viola-Jones method, (b) Hyun [9], (c) Choi [8], and (d) proposed method

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

Fig. 4. Experimental result using the ‘Solvay Conference 1927’ image: (a) traditional Viola-Jones method, (b) Hyun [9], (c) Choi [8], and (d) proposed method

proposed by Hyun [9] necessitates a different number of classification iterations. However, the method proposed by Choi [8] significantly reduces the number of iterations by 3.45M. Notably, our proposed method demonstrates a further reduction, requiring 2.37M iterations. This represents an 84.7272%, 84.7058%, and 31.3778% reduction compared to conventional methods. For the ‘Solvay Conference 1927’ image, conventional methods demand 73.465M, 73.095M, and 16.481M classification iterations, while our proposed method requires 10.444M, achieving an 85.7838%, 85.7118%, and 36.6306% reduction compared to conventional methods.

In conclusion, the experimental results across the three images consistently demonstrate that our proposed method requires the fewest number of classification iterations. This implies that the proposed method exhibits a faster processing speed compared to conventional methods.

TABLE I
THE NUMBER OF CLASSIFICATION ITERATION FOR TWO TEST IMAGES WHEN USING THE PROPOSED AND CONVENTIONAL METHODS

Image	Works			
	Traditional	Hyun [9]	Choi [8]	Our
Lena	15,502,079	15,480,375	3,450,202	2,367,604
Solvay Conference 1927	73,465,180	73,094,971	16,481,076	10,443,967

IV. Conclusion

In object detection algorithms utilizing the Viola-Jones method, the processing time tends to increase with the number of classification iterations. To mitigate this challenge, this paper introduces a swift object detection algorithm that integrates an edge-based operation skip scheme and edge component calibration into the Viola-Jones method. The experimental results highlight the efficacy of our proposed method, showcasing the lowest number of classification iterations. Moreover, the introduction of the refined edge component calibration method ensures a reduction in precision while maintaining recall compared to conventional methods. As a result, our proposed method is proven to be well-suited for object detection systems requiring rapid processing speeds.

In future work, we aim to implement the object detection systems on the FPGA platform. Additionally, further experiments on diverse test images will be conducted to comprehensively assess detection performance and processing speed.

References

[1] A. Mishra, S. Lee, D. Kim, and S. Kim, “In-cabin monitoring system for autonomous vehicles,” *Sensors*, vol. 22, no. 12, p. 4360, 2022.

[2] J. Mao, S. Shi, X. Wang, and H. Li, “3d object detection for autonomous driving: A comprehensive survey,” *International Journal of Computer Vision*, pp. 1–55, 2023.

[3] M. Da'San, A. Alqudah, and O. Debeir, "Face detection using viola and jones method and neural networks," in *2015 International Conference on Information and Communication Technology Research (ICTRC)*. IEEE, 2015, pp. 40–43.

[4] A. Srivastava, S. Mane, A. Shah, N. Shrivastava, and B. Thakare, "A survey of face detection algorithms," in *2017 International Conference on Inventive Systems and Control (ICISC)*. IEEE, 2017, pp. 1–4.

[5] R. Vij and B. Kaushik, "A survey on various face detecting and tracking techniques in video sequences," in *2019 International conference on intelligent computing and control systems (ICCS)*. IEEE, 2019, pp. 69–73.

[6] 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. CVPR 2001*, vol. 1. Ieee, 2001, pp. I–I.

[7] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, pp. 137–154, 2004.

[8] C.-H. Choi, J. Kim, J. Hyun, Y. Kim, and B. Moon, "Face detection using haar cascade classifiers based on vertical component calibration," *Human-centric Computing and Information Sciences*, vol. 12, no. 11, 2022.

[9] J. Hyun, J. Kim, C.-H. Choi, and B. Moon, "Hardware architecture of a haar classifier based face detection system using a skip scheme," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2021, pp. 1–4.